EPIDEMIOLOGY

Reproducibility of non-specific bronchial challenge in adults: implications for design, analysis and interpretation of clinical and epidemiological studies

S Chinn, J P Schouten

Thorax 2005;**60**:395–400. doi: 10.1136/thx.2004.039230

See end of article for authors' affiliations

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

Received 13 December 2004
Accepted 9 March 2005

Background: Poor reproducibility of an outcome measure reduces power and, in an independent variable, biases results. The intraclass correlation coefficient measures loss of power and degree of bias. Information is lacking on the intraclass correlation coefficient for bronchial responsiveness and factors affecting reproducibility.

Methods: Papers containing information on reproducibility of bronchial responsiveness were identified using a Medline search and citations. Within and between person components of variance of PD_{20} or PC_{20} were expressed in doubling dose or concentration units, and the intraclass correlation coefficient calculated when not reported.

Results: Results were extracted from 32 papers. Intraclass correlation coefficients were over 0.9 in short term studies of highly selected asthmatic patients, but larger and most long term studies had lower intraclass correlation coefficients, less than 0.5 in some cases, due to greater within person or lower between person variation. Reproducibility of dose or concentration-response slope was generally higher, but still less than that of forced expiratory volume in 1 second.

Conclusions: Information is available to calculate sample size for studies with bronchial responsiveness as the outcome, but results when bronchial responsiveness is an explanatory variable may be misleading.

Clinical assessment of change in a patient should properly be made with reference to variation in change in healthy subjects.[1] This variation is a combination of measurement error and true within person fluctuation. The latter component is likely to increase the longer the time interval between assessments. Reviews of short term repeatability of bronchial responsiveness (BHR) reported the width of the 95% range for a single measurement of an individual's BHR, expressed as the dose (PD_{20}) or concentration (PC_{20}) estimated to produce a 20% fall in forced expiratory volume in 1 second (FEV_1), from under 1 doubling dose or concentration to over 2.5.[2 3] A 95% range for change is obtained by multiplying the within person standard deviation by $\sqrt{2}$. Despite the fact that this gives a 95% range that may be over 3.5 doubling doses in width, authors have considered that, in the short term, ''test results vary little''.[4] Limited information on longer term reproducibility has been reported,[2] the width of the 95% range for a single PC_{20} being up to 4 doubling concentrations.

A clinician needs to know not only how variable a single reading may be for an individual patient, but how the variation compares with that between patients. The greater within patient variation is in relation to between patient variation, the less information a single measurement provides for the individual. In epidemiological studies poor repeatability of a continuous outcome variable reduces power to detect relations with risk factors, as the total variance is increased. The loss of power is a function of the ratio of the total variance including error to the true variance without error, where the term ''error'' is used to encompass both measurement error and true within person fluctuation over the relevant time period. Hence, for the clinician and the epidemiologist, it is useful to know the ratio of true between person variation to the total variance. This quantity is known as the intraclass correlation coefficient (ICC). It is a dimensionless quantity that takes the value 1.0 when there is no error and 0.0 when there is no true variation. An increase in measurement error decreases its value, but selection of a sample of homogenous individuals reduces true variation and hence also decreases the ICC.

Poor repeatability of an explanatory variable leads to bias in effect estimates. The bias, known formerly as attenuation by statisticians[5] but by epidemiologists as regression dilution,[6] is also a function of the ICC which is often referred to as ''reliability''.[6] Error in the independent variable in a regression analysis produces a regression line that is less steep than the true relation. When the relative sizes of the variance due to ''error'' and the total variation are known, the bias can be determined.[7] The true value of the regression coefficient can be estimated as observed value/ICC in the absence of covariates, although bias is difficult to predict in multiple regression.[8] Hence, despite the ICC being criticised as a measure of agreement,[9] it is clearly of great importance as a measure of repeatability.

Very limited information on ICCs has been given previously.[3] This paper reports a review of the literature for estimates of repeatability of BHR, and of ICCs in particular, and assesses the implications of findings for analysis and interpretation of BHR.

METHODS

A Medline search of the subject heading ''bronchial hyperreactivity'' or one of its synonyms as a keyword (airway/bronchial (hyper)reactivity/responsiveness), combined with ''reproducibility of results'', or ICC, repeatability or reliability as a keyword, was used to obtain abstracts of potential papers. Papers were additionally identified from reference lists. Papers were limited to studies of adults and English language publication. Abstracts were read to exclude studies that did not have repeat measurement of BHR, or where change in BHR with change in treatment or other conditions was the subject of study. Provocation agents were

limited to histamine and methacholine—that is, a small number of papers on reproducibility of BHR to carbachol, cold air, exercise or hypertonic saline were excluded. Measurements repeated on the same day were not included, and where methods of administration were compared in the same subjects only the preferred method was included.

Repeatability data were extracted and within and between subject components of variation were expressed in doubling dose or concentration standard deviations. Papers which did not report the within subject standard deviation or the ICC or allow either to be estimated were omitted. Unless otherwise stated, a published 95% range for a single value was assumed to be calculated as ± 2 within subject standard deviations. Where limits were stated to be a ''confidence interval'', statistics were derived only if it was clear from the text whether the limits were calculated from a standard deviation (that is, a 95% range) or from a standard error (that is, a true confidence interval). When data were presented only graphically they were measured from the graph, taking account of differing scales on the axes where necessary. Raw data were used if given and analysed by one way analysis of variance of dose or concentration in doubling dose units by subject, and components of variance calculated,[7] from which the ICC was derived. Based on the distribution of length of follow up in the papers, an arbitrary division into short term and long term follow up was made at a cut off of 4 months.

RESULTS

The Medline search produced 101 abstracts, of which 37 potentially met the inclusion criteria and 23 were found to have useful repeatability data.[4 10–31] Of the 14 exclusions, eight were found not to meet the inclusion criteria on reading the full paper, one gave data for a subset of data reported in another paper, and five did not report results in a form that allowed derivation of components of variance or the ICC. A further eight papers were identified from citations[32–39] and one study primarily of other measures was included.[40] Where only a measure of within person variation, or only the ICC, was stated but data were represented graphically, there was good agreement between the stated estimate and the corresponding value calculated from the measured data except in one case mentioned below.

Short term repeatability

Table 1 gives short term estimates of repeatability of PD_{20} or PC_{20} from eight studies published before 1987. These were each carried out on a small number of asthmatic patients. ICCs were above 0.9 when the within person standard deviation was less than 0.5 doubling doses, and 0.97 or more when combined with a between person standard deviation of at least 2.0 doubling doses. In one study the difference between the stated within person variation (1.0) and that derived from the graphical data (0.7) was noteworthy.[38]

Table 2 shows corresponding estimates from nine studies published from 1987 to 1991. Each of these gave a measure of within person variability, and most the ICC as well. Two of the studies were on population samples,[13 14] but the larger study mostly comprised participants who had a measurable PD_{20} at the first occasion,[13] and the smaller recruited participants with wheeze or asthma.[14] Three studies achieved low within person variation[15 16 40] but most studies had greater within person variation than the earlier studies, and hence lower ICCs. The study of hospital personnel had low between person variation and hence a low ICC. Table 3 shows estimates published from 1993 to 2001. All but one of these studies was carried out in asthmatic patients.

Long term repeatability

Long term repeatability over a period of 4 months or more was estimated in seven studies (table 4). Three of these studies were general population studies which gave lower ICCs than other studies. The largest study found an ICC for PD_{10} of 0.32 for asymptomatic and 0.42 for symptomatic subjects.[4] New results including an extra follow up survey gave an overall ICC of 0.37, with a within person standard deviation of 1.0 doubling concentrations that was comparable to other studies, but lower between person variation. An ICC of 0.45 for PD_{20} was obtained by Beckett *et al*,[30] with the largest within person standard deviation of any study. The third study, carried out in general practice, found an ICC of 0.48 for 27 subjects with complete data and measurable PD_{20} on six occasions, but higher ICCs (0.56 and 0.68) when all first year and all second year pairs were analysed.[24] The three long term studies on selected asthmatic subjects had lower within person standard deviations and higher ICCs.[10 28 37] The study of aluminium smelter workers included data only for 36 people, those with a 20% fall in FEV_1 by the maximum dose of 6.14 μmol at each occasion, in the calculation of ICC for PD_{20}.[39] New results from the Vlagtwedden/Vlaardingen study showed increasing within person variation, and hence decreasing ICC, with increasing length of follow up (not shown).

Repeatability of dose-response slope

Table 5 shows estimates of short term repeatability of the FEV_1-dose response slope from two studies and long term

Table 1 Short term (less than 4 months) repeatability of PD_{20} or PC_{20} in early studies on asthmatic patients and normal controls. Measurements in duplicate except where otherwise stated

Year and source	No of participants	Age range (years)	Provocation agent and maximum concentration/ cumulative dose	Time interval	Summary statistic	Within person SD (doubling doses or concentrations)	Between person SD (doubling doses or concentrations)	Intraclass correlation coefficient (ICC)
1978[32]	11 asthmatic, 3 normal	Unknown	Histamine, 16 mg/ml	Within 1 week	PC_{20}-FEV_1	0.2*	2.5*	0.994
1978[32]	11 asthmatic, 2 normal	Unknown	Methacholine, 16 mg/ml	Within 1 week	PC_{20}-FEV_1	0.3*	2.6*	0.990
1981[33]	10 asthmatic	16–65	Histamine, 16 mg/ml	Within 2 weeks	PC_{20}-FEV_1	0.3	2.0	0.97
1981[34]	12 asthmatic, studied 4 times	25–63	Histamine, 11.5 mg/ml	1–12 days	PC_{20}-FEV_1	0.4	1.5	0.94
1983[11]	15 mixed, all atopic	Unknown	Histamine, 7.8 μmol	Within 10 days	PD_{20}-FEV_1	0.4*	1.5*	0.93*
1983[35]	20 asthmatic	Unknown	Histamine, 8 mg/ml	Within 5 days	PC_{20}-FEV_1	0.3*	2.0*	0.98*
1983[36]	18 asthmatic	19–55	Histamine, 32 mg/ml	Within 2 weeks	PC_{20}-FEV_1	0.8	2.2, calculated from ICC, and within person SD	0.88
1985[38]	27 mixed	17–49	Histamine, 16 mg/ml	65 days	PC_{20}-FEV_1	1.0 0.7*	2.3*	0.91*
1986[12]	24 asthmatic	18–55	Histamine, 7.8 μmol	1–7 days	PD_{20}-FEV_1	1.1	1.7*	0.72*

*Estimated from graphical information.
SD, standard deviation.

Table 2 Short term (less than 4 months) repeatability of PD_{20} or PC_{20} in studies published from 1987 to 1991

Year and source	Participants	Age range (years)	Provocation agent and maximum concentration/ cumulative dose	Time interval	Summary statistic	Within person SD (doubling doses or concentrations)	Between person SD (doubling doses or concentrations)	Intraclass correlation coefficient (ICC)
1987[13]	90 non-random population sample	18–64	Histamine, 4 μmol	1 –14 days	PD_{20}-FEV_1	0.9	1.9	0.81
1988[14]	25 with wheeze or asthma, population sample	18–75	Histamine, 4 μmol	Few weeks	PD_{20}-FEV_1	1.2	1.3*	0.57*
1988[14]	27 with wheeze or asthma, population sample	18–75	Methacholine, 12 μmol	Few weeks	PD_{20}-FEV_1	1.0	1.9*	0.84*
1988[15]	20 asthmatic patients	65–82	Methacholine, 32.65 μmol	Within 10 days	PD_{20}-FEV_1	0.5	2.7*	0.97*
1988[16]	20 asthmatic patients	19–61	Methacholine, 32.65 μmol	1–10 days	PD_{20}-FEV_1	0.3	2.4	0.98
1988[16]	20 asthmatic patients	19–61	Histamine, 10.4 μmol	1–10 days	PD_{20}-FEV_1	0.3	2.6	0.99
1989[17]	20 asthmatic patients	19–66	Methacholine, 24.5 μmol	1–7 days	PD_{20}-FEV_1	1.0	1.9	0.79
1989[40]	19 hospital personnel	23–38	Histamine, 6.83 μmol	1 week	PD_{20}-FEV_1	0.4	0.4	0.42
1990[18]	10 asthmatic patients	18–55	Methacholine, 16 mg/ml	Daily for 5 days	PC_{20}-FEV_1	1.2	1.6	0.64
1991[19]	14 healthy ''responsive''	Mean (SD) 22.2 (3.2)	Methacholine, 85.2 μmol	1–14 days	PD_{20}-FEV_1			0.89
1991[20]	20 asthmatic, experienced	20–59	Methacholine, 25 μmol	Within 2 weeks	PD_{20}-FEV_1	0.7	1.9*	0.91*
1991[20]	20 asthmatic, inexperienced	20–59	Methacholine, 25 μmol	Within 2 weeks	PD_{20}-FEV_1	0.9	1.7*	0.76*

*Estimated from graphical information.
SD, standard deviation.

ICC from four. Dose or concentration-response slope was calculated from two data points,[41] except for one study which used regression of percentage decline in FEV_1 on dose.[23] This study reported data from 104 participants which included 90 whose repeatability of PD_{20} was given in table 2.[13] The latter only included people with a measurable PD_{20} on at least one occasion, while the FEV_1-dose response slope was calculated for each participant who received two or more doses of histamine. The ICCs were 0.89 for slope and 0.81 for PD_{20}. The study of aluminium smelter workers, which included data only for persons with two measurable PD_{20} values in the PD_{20}-ICC, found a much higher ICC for log dose-response slope (0.73 compared with 0.28).[39] Trigg *et al*[24] found a higher ICC for the dose-response slope (0.75) than for PD_{20} (0.48), and Beckett *et al*[30] slightly higher (0.54 compared with 0.45).

DISCUSSION

Variation in ICC

The early studies on short term repeatability in selected asthmatic patients achieved good repeatability, as indicated by the within person standard deviation of less than 0.5 doubling doses or concentrations and hence a high ICC. Early enthusiastic exponents of bronchial challenge may have taken greater care over procedures or selected highly cooperative patients. Many later studies, particularly the larger population studies, had a within person standard deviation of around 1.0 doubling doses or concentrations. Selection of subjects determines the between subject variation. A population study has a large majority of ''non-responsive'' individuals whose values are clustered at the maximum dose or concentration; even when this is high, the use of a logarithmic scale reduces the apparent variation at the upper end of the scale.

Variation is expressed on the doubling dose or concentration scale as this is the most appropriate for PD_{20} or PC_{20},[42] but ICCs are independent of linear transformation—that is, they are the same on any logarithmic scale. Repeatability of histamine and methacholine BHR appears similar on a logarithmic scale. There is no agreement over scale for the dose-response slope (table 5) but correlation with log PD_{20} has been shown to be high when the dose-response slope is reciprocally transformed[23] or log-transformed.[24]

The Pearson correlation coefficient measures the degree of any linear relation between two variables. It is therefore inappropriate for repeatability studies as, for example, change in mean BHR over time would not affect it but does lower the ICC. The unsuitability of the Pearson correlation coefficient for method comparison and repeatability was made clear in 1986[43] but, despite this, several later papers reported it, although within person variation was generally also reported.

There were too few long term studies with fixed follow up time to relate within person variation to length of follow up. Unpublished results from the Vlagtwedden/Vlaardingen study suggest an increase in within person variation with length of follow up. It is unclear whether the lower ICCs in

Table 3 Short term (less than 4 months) repeatability of BHR in studies published from 1993 to 2001

Year and source	Participants	Age range (years)	Provocation agent and maximum concentration/ cumulative dose	Time interval	Summary statistic	Within person SD (doubling doses or concentrations)	Between person SD (doubling doses or concentrations)	Intraclass correlation coefficient (ICC)
1993[21]	14 asthmatic patients	19–65	Histamine, 6.9 μmol	1 day	PD_{15}-FEV_1	0.4*	1.8*	0.95
1993[22]	20 asthmatic patients	18–65	Methacholine, 32.65 μmol	Within 10 days	PD_{20}-FEV_1	0.6	2.0*	0.92*
1995[26]	19 asthmatic patients	18–45	Methacholine, 32.65 μmol	1 day	log PD_{20}-FEV_1	0.8*	2.7*	0.92*
1995[27]	10 non-smoking asthmatic patients	18–57	Histamine, 1.4 μmol	2–7 days	log PD_{20}-FEV_1	0.6	1.1	0.81
1996[29]	11 asthmatic ''hyperresponsive'' patients	19–62	Methacholine, 35.2 μmol	1–7 days	log PD_{20}-FEV_1	0.4	1.1*	0.87*
2001[31]	34 non-asthmatic	17–78	Methacholine, 64 mg/ml	Within 3 weeks	log PC_{20}-FEV_1	0.7*	1.4*	0.71*

*Estimated from graphical information.
SD, standard deviation.

Table 4 Long term (more than 4 months) repeatability of BHR

Year and source	Participants	Age range (years)	Provocation agent and maximum concentration/ cumulative dose	Time interval	Summary statistic	Within person SD (doubling doses or concentrations)	Between person SD (doubling doses or concentrations)	Intraclass correlation coefficient (ICC)
1982[10]	35 asthmatic	Not stated	Histamine, 16 mg/ml	10–30 months	PC_{20}-FEV_1	1.0*	1.7*	0.74*
1983[37]	10 atopic asthmatic patients	23–41 at baseline	Histamine, 8.0 mg/ml	14–28 months	PC_{20}-FEV_1	0.5*	1.4*	0.87
1990[39]	36 ''responsive'' workers	Not stated	Methacholine, 6.14 μmol	1 year	PD_{20}-FEV_1			0.28
1993[4]	1413 asymptomatic population sample	15–54 at baseline	Histamine, 32 mg/ml	3–22 years	PC_{10}-FEV_1			0.32
1993[4]	803 symptomatic population sample	15–54 at baseline	Histamine, 32 mg/ml	3–22 years	PC_{10}-FEV_1			0.42
V/V	2173 population sample	15–54 at baseline	Histamine, 32 mg/ml	3–22 years	PC_{10}-FEV_1	1.0	0.8	0.37
1994[24]	27 general practice, $PD_{20} \leqslant 247$ μmol on 6 occasions	18–75	Methacholine, 247 μmol	4–24 months	PD_{20}-FEV_1	1.1	1.1, calculated from ICC and within person SD	0.48
1996[28]	10 asthmatic, PC_{20} <9 mg/ml	25–82	Histamine, 16 mg/ml	6 months	PC_{20}-FEV_1	0.7*	1.4*	0.80*
1997[30]	88 healthy working adults	Not stated	Methacholine, 13.06 μmol (in repeatability data)	1–3 years	PD_{20}-FEV_1	1.8	1.6	0.45

*Estimated from graphical information.
SD, standard deviation.
V/V unpublished results from Vlagtwedden/Vlaardingen study; published results also included.[4]

table 4 compared with tables 1–3 were primarily due to longer follow up with increased within person standard deviation, or to sampling from a general population and lower between person variation. The one population study of short term repeatability had a within person standard deviation in line with short term studies of asthmatic patients[13 23] but also with the long term Vlagtwedden/ Vlaardingen study (table 4), while the long term population study of Beckett *et al* had greater within person variation.[30] Hence, the effect of selection of subjects on between person variation and the effect of length of follow up on within person variation both influence the ICC. However, restriction of data to participants with two measurable PD_{20} values decreases between person variation and hence the ICC.[23] The dose-response slope, which can be estimated for people who do not have a measurable PD_{20}, had a greater ICC than PD_{20} in one study[24] and slightly larger in two others.[23 30]

Many of the estimates of ICC are based on a small sample but, due to the many reasons for variability in the components of variation in BHR and hence in ICC, it would not be appropriate to pool estimates. For this reason, no attempt was made to carry out a fully systematic review. Confidence intervals for the ICC are wide; for example, Seppälä gave a 95% confidence interval of 0.70 to 0.96 for an ICC of 0.89 found for $\ln(PD_{20})$ in 14 responsive healthy subjects.[19]

BHR as outcome variable in a cross sectional study

In carefully controlled studies with selected participants an ICC of 0.99 can be achieved, as high as that for FEV_1.[44] However, such a high ICC is unlikely to be achieved in larger studies. In the studies which assessed repeatability of FEV_1 and BHR in the same subjects, the ICC for BHR was lower than that for FEV_1[13 18 30 34 40] so that studies of BHR generally require more participants than those on FEV_1 to detect an equivalent size of effect. The standard deviation that should be used in a sample size calculation is the total short term variation in a study with similar participants; this can be calculated by adding the squares of the within and between standard deviation and taking the square root of the result.

Change in BHR as outcome in short term follow up studies

The standard deviation of change in any continuous outcome is calculated by multiplying the within person standard deviation by the square root of two. Appropriate within person standard deviations in tables 1–3 can therefore be used to calculate sample size or power. In randomised controlled trials the recommended analysis is of final outcome with the baseline value as a covariate,[45] as this increases power and is unbiased as baseline mean values will be equal on average. However, this method is inadvisable in an observational study. The regression coefficient of final on initial value is biased towards zero. It is used to adjust the estimated means at follow up of groups that differ in mean value at baseline and so will affect the comparison of interest and can even reverse the sign of the difference.[5] In addition, Schouten and Tager[46] have explained why adjusting for baseline may give misleading results. The analysis of final outcome with baseline as covariate has little to recommend it

Table 5 Intraclass correlation coefficients for measures of FEV_1-dose-response slope

Year and source	Participants	Age range (years)	Provocation agent and maximum concentration/cumulative dose	Time interval	Summary statistic	Intraclass correlation coefficient (ICC)
1990[39]	726 workers	Not stated	Methacholine, 6.14 μmol	1 year	Log dose-response slope	0.73
1991[19]	14 healthy ''responsive''	Mean (SD) 22.2 (3.2)	Methacholine, 85.2 μmol	1–14 days	Log dose-response slope	0.99
1991[19]	16 healthy ''non-responsive''	Mean (SD) 25.4 (3.3)	Methacholine, 85.2 μmol	1–14 days	Log dose-response slope	0.50
1993[23]	104 population sample	18–64	Histamine, 4 μmol	1–14 days	1/(dose response slope +10)	0.89
1994[24]	67 general practice on 6 occasions	18–75	Methacholine, 247 μmol	4–24 months	Log dose-response slope	0.75
1994[25]	12 healthy non-smokers	Mean (SD) 27.4 (3.4)	Histamine, 16 mg/ml	1–3 days	Dose-response slope	0.97
1997[30]	88 healthy working adults	Not stated	Methacholine, 13.06 μmol (in repeatability data)	1–3 years	Log dose-response slope	0.54

FEV_1, forced expiratory volume in 1 second.

in non-randomised studies and, for an outcome with an ICC that may be as low as 0.5 in some circumstances, it is definitely to be avoided in such studies.

Change in BHR as outcome in long term follow up studies

It is likely that the change in BHR over several months or years will be more variable than in the short term. Although the number of studies is small with most of the information from population based studies, lower ICCs are unlikely to be due wholly to differences in participants. Firstly, the short term population study found a relatively high ICC due to low within person variation[13 23] and, secondly, the large long term study found even lower ICCs on adjustment for individual explanatory variables, as between person variation was reduced proportionally more than within person variation.[4] The within person standard deviations in table 4 can be used in sample size calculations, although they will be conservative as some of the within person variation will be explained by changes in explanatory variables. On the other hand, the use of standard deviations in tables 1–3 may result in too small a sample size. The recommendation to analyse absolute change, and not final adjusted for initial value, applies even more strongly to long term than to short term observational studies.

BHR as an independent variable

A number of authors have used BHR as an independent variable, particularly as a predictor of decline in FEV_1,[47–49] dividing participants into ''responders'' and ''non-responders''. Few authors have reported a kappa statistic for repeatability of dichotomised BHR, but it can be expected to be similar in value to the ICC. BHR has a unimodal continuous distribution in the general population[50 51] and is not a fixed state, as many authors seem to assume. The problem—whether BHR is dichotomised or not—is the same as that of using baseline BHR as a covariate when final BHR is the outcome in a longitudinal study, that there will be bias in the regression coefficient of outcome on BHR and also of the other regression coefficients in a multiple regression. Correction for bias requires estimates of the ICC for variances and covariances of the explanatory variables[6 8] which can only be determined from a repeatability study of all covariates subject to within person variability carried out on all, or a substantial random sample, of the participants unless certain assumptions are met.[52]

Conclusion

The analysis of BHR as an outcome variable is straightforward and there is considerable information to allow studies to be planned with adequate sample size to take account of the inherent variation. PD_{20} or PC_{20} are known only to be above the maximum dose or concentration (that is, ''censored'') when a 20% fall in FEV_1 has not occurred when the challenge is stopped. This has often led authors to express BHR as ''responsive'' or ''not responsive'' and to use logistic regression to analyse the data, but greater power is achieved if regression methods for censored data are used or a dose-response slope or other continuous outcome analysed.[53] This is reinforced by ICCs for the dose-response slope being at least as high and probably greater than those for PD_{20}.

In contrast, analysis of BHR as an explanatory variable is liable to give biased and possibly misleading results. This is true of any explanatory variable for which the short term ICC may be as low as 0.5. BHR contrasts with FEV_1 as the short term ICC for FEV_1 can be presumed to be over 0.9[13 18 34 40] and has been reported to be 0.89 over 1–3 years.[30] Lung function has been shown to be strongly associated with BHR in cross sectional studies, part of which may due to inherent dependence of BHR summary statistics on FEV_1. Analyses of BHR as the outcome therefore need to adjust for lung function even if a causal role is not assumed.

Rijcken and Weiss posed the question of whether a lower level of FEV_1 is a cause or a result of increased airway responsiveness and stated that longitudinal analyses are necessary to answer the question.[54] We can add to this that either multiple measurements of BHR should be made to increase precision[4] or the regression coefficients should be adjusted for lack of repeatability. Unfortunately, if ICCs of variances are highly variable, those of covariances may be even harder to estimate and extrapolation from another study is unlikely to be sound. Unless researchers take steps to increase precision, the inclusion of BHR as an explanatory variable may be misleading.

Authors' affiliations

S Chinn, Department of Public Health Sciences, King's College London, London SE1 3QD, UK
J P Schouten, Department of Epidemiology and Bioinformatics, University Medical Center, 9700 RB Groningen, The Netherlands

No competing interests declared.

REFERENCES

1 **Harris EK**, Boyd JC. *Statistical bases of reference values in laboratory medicine*. New York: Marcel Dekker, 1995:223.
2 **Weeke B**, Madsen F, Frølund L. Reproducibility of challenge tests at different times. *Chest* 1987;**91**:83–9S.
3 **Chinn S**, Sunyer J. Bronchial hyperresponsiveness. In: Annesi-Maesano I, Gulsvik A, Viegi G, eds. *Respiratory epidemiology in Europe*. Sheffield: European Respiratory Society, 2000:199–215.
4 **Rijcken B**, Schouten JP, Weiss ST, *et al*. Long-term variability of bronchial responsiveness to histamine in a random population sample of adults. *Am Rev Respir Dis* 1993;**148**:944–9.
5 **Goldstein H**. *The design and analysis of longitudinal studies*. London: Academic Press, 1979:133–4.
6 **Dyer AR**, Elliott P, Shipley M for the INTERSALT Cooperative Research group. Urinary electrolyte excretion in 24 hours and blood pressure in the INTERSALT study. *Am J Epidemiol* 1994;**139**:940–51.
7 **Armitage P**, Berry G, Matthews JNS. *Statistical methods in medical research*. 4th edn. London: Blackwell Science, 2002.
8 **Cook NR**, Kumanika SK, Cutler JA. Effect of change in sodium excretion on change in blood pressure corrected for measurement error. *Am J Epidemiol* 1998;**148**:431–44.
9 **Bland JM**, Altman DG. A note on the use of the intraclass correlation coefficient in the evaluation of agreement between two methods of measurement. *Comput Biol Med* 1990;**20**:337–40.
10 **Juniper EF**, Frith PA, Hargreave FE. Long-term stability of bronchial responsiveness to histamine. *Thorax* 1982;**37**:288–91.
11 **Yan K**, Salome C, Woolcock AJ. Rapid method for measurement of bronchial responsiveness. *Thorax* 1983;**38**:760–5.
12 **Britton J**, Mortagy A, Tattersfield A. Histamine challenge testing: comparison of three methods. *Thorax* 1986;**41**:128–32.
13 **Chinn S**, Britton JR, Burney PG, *et al*. Estimation and repeatability of the response to inhaled histamine in a community survey. *Thorax* 1987;**42**:45–52.
14 **Higgins BG**, Britton JR, Chinn S, *et al*. Comparison of histamine and methacholine for use in bronchial challenge tests in community studies. *Thorax* 1988;**43**:605–10.
15 **Connolly MJ**, Kelly C, Walters EH, *et al*. An assessment of methacholine inhalation tests in elderly asthmatics. *Age Ageing* 1988;**17**:123–8.
16 **Connolly MJ**, Avery AJ, Walters EH, *et al*. The relationship between bronchial responsiveness to methacholine and bronchial responsiveness to histamine in asthmatic subjects. *Pulmon Pharmacol* 1988;**1**:53–8.
17 **Knox AJ**, Coleman HE, Britton JR, *et al*. A comparison of three measures of the response to inhaled methacholine. *Eur Respir J* 1989;**2**:736–40.
18 **Trigg CJ**, Jhalli N, Herdman MJ, *et al*. The daily variability of bronchial responsiveness to methacholine. *Eur Respir J* 1990;**3**:867–71.
19 **Seppälä O-P**. The dose-response slope: a useful method for expressing the results of methacholine provocation tests in healthy subjects? *Respir Med* 1991;**85**:365–71.
20 **Knox AJ**, Wisniewski A, Cooper S, *et al*. A comparison of the Yan and a dosimeter method for methacholine challenge in experienced and inexperienced subjects. *Eur Respir J* 1991;**4**:497–502.
21 **Sovijärvi ARA**, Malmberg LP, Reinikainen K, *et al*. A rapid dosimetric method with controlled tidal breathing for histamine challenge. *Chest* 1993;**104**:164–70.
22 **Beach JR**, Young CL, Avery AJ, *et al*. Measurement of airway responsiveness to methacholine: relative importance of the precision of drug delivery and the method of assessing response. *Thorax* 1993;**48**:239–43.

23 **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.
24 **Trigg CJ**, Tooley M, D'Souza MF, *et al*. Factors affecting the long-term variability of bronchial responsiveness in an adult general practice population. *Eur Respir J* 1994;**7**:703–9.
25 **Seppälä O-P**. A method for measuring the effects of anticholinergics on histamine-induced bronchoconstriction in normal subjects. Oxitropium bromide provides dose-dependent protection. *Respir Med* 1994;**88**:273–29.
26 **Beach JR**, Stenton SC, Connolly MJ, *et al*. Effects of diurnal variation and prolonged refractoriness on repeated measurements of airways responsiveness to methacholine. *Thorax* 1995;**50**:235–9.
27 **Wood-Baker R**, Town GI, Benning B, *et al*. The reproducibility and effect on non-specific airway responsiveness of inhaled prostaglandin D_2 and leukotriene D_4 in asthmatic subjects. *Br J Clin Phamacol* 1995;**39**:119–23.
28 **Gibbons WJ**, Sharma A, Lougheed D, *et al*. Detection of excessive bronchoconstriction in asthma. *Am J Respir Crit Care Med* 1996;**153**:582–9.
29 **Hedman J**, Alanko K, Nieminen MM. Repeatability of a rapid dosimetric method for methacholine challenge using a pocket turbine spirometer for FEV_1 measurements. *Clin Physiol* 1996;**16**:353–9.
30 **Beckett WS**, Pace PA, Sferlazza SJ, *et al*. Annual variability in methacholine responsiveness in nonasthmatic working adults. *Eur Respir J* 1997;**10**:2515–21.
31 **Sundblad B-M**, Malmberg P, Larsson K. Comparison of airway conductance and FEV_1 as measures of airway responsiveness to methacholine. *Clin Physiol* 2001;**21**:673–81.
32 **Juniper EF**, Frith PA, Dunnett C, *et al*. Reproducibility and comparison of responses to inhaled histamine and methacholine. *Thorax* 1978;**33**:705–10.
33 **Ryan G**, Dolovich MB, Roberts RS, *et al*. Standardization of inhalation provocation tests: two techniques of aerosol generation and inhalation compared. *Am Rev Respir Dis* 1981;**123**:195–9.
34 **Ruffin RE**, Alpers JH, Crockett AJ, *et al*. Repeated histamine inhalation tests in asthmatic patients. *J Allergy Clin Immunol* 1981;**67**:285–9.
35 **Cockcroft DW**, Berscheid BA, Murdock KY. Measurement of responsiveness to inhaled histamine using FEV_1: comparison of PC_{20} and threshold. *Thorax* 1983;**38**:523–6.
36 **Dehaut P**, Rachielle A, Martin RR, *et al*. Histamine dose-response curves in asthma: reproducibility and sensitivity of different indices to assess response. *Thorax* 1983;**38**:516–22.
37 **Löwhagan O**, Lindholm NB. Short-term and long-term variation in bronchial response to histamine in asthmatic patients. *Eur J Respir Dis* 1983;**64**:466–72.
38 **Madsen F**, Rathlou NH, Frolund L, *et al*. Short and long term reproducibility of responsiveness to inhaled histamine: R_t compared to FEV_1 as measurement of response to challenge. *Eur J Respir Dis* 1985;**67**:193–203.
39 **Abramson MJ**, Saunders NA, Hensley MJ. Analysis of bronchial reactivity in epidemiological studies. *Thorax* 1990;**45**:924–9.
40 **Neild JE**, Twort CHC, Chinn S, *et al*. The repeatability and validity of respiratory resistance measured by the forced oscillation technique. *Respir Med* 1989;**83**:111–8.
41 **O'Connor G**, Sparrow D, Taylor D, *et al*. Analysis of dose-response curves to methacholine. *Am Rev Respir Dis* 1987;**136**:1412–17.
42 **Peat JK**, Unger WR, Combe D. Measuring change in logarithmic data, with special reference to bronchial responsiveness. *J Clin Epidemiol* 1994;**47**:1099–108.
43 **Bland JM**, Altman DG. Statistical methods for assessing agreement between two methods of clinical measurement. *Lancet* 1986;**i**:307–10.
44 **Oldham PD**, Cole TJ. Estimation of the FEV_1. *Thorax* 1983;**38**:662–7.
45 **Senn SJ**. Baseline adjustment in longitudinal studies. In: Armitage P, Colton T, eds. *Encyclopedia in biostatistics*. Volume 1. Chichester: John Wiley, 1998:253–7.
46 **Schouten JP**, Tager IB. Interpretation of longitudinal studies. *Am J Respir Crit Care Med* 1996;**154**:S278–84.
47 **van Schayk CP**, Dompeling E, van Herwaarden CLA, *et al*. Interacting effects of atopy and bronchial hyperresponsiveness on the annual decline in lung function and the exacerbation rate in asthma. *Am Rev Respir Dis* 1991;**144**:1297–301.
48 **Frew AJ**, Kennedy SM, Chan-Yeung M. Methacholine responsiveness, smoking, and atop as risk factors for accelerated FEV_1 decline in male working populations. *Am Rev Respir Dis* 1992;**146**:878–83.
49 **Rijcken B**, Schouten JP, Xu X, *et al*. Airway hyerresponsiveness to histamine associated with accelerated decline in FEV_1. *Am J Respir Crit Care Med* 1995;**151**:1377–82.
50 **Cockcroft DW**, Bersheid BA, Murdock KY. Unimodal distribution of bronchial responsiveness to inhaled histamine in a random human population. *Chest* 1983;**83**:751–4.
51 **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.
52 **Tosteson TD**, Buonaccorsi JP, Demidenko E. Covariate measurement error and the estimation of random effect parameters in a mixed model for longitudinal data. *Stat Med* 1998;**17**:1959–71.
53 **Chinn S**. Methodology of bronchial responsiveness. *Thorax* 1998;**53**:984–8.
54 **Rijcken B**, Weiss ST. Longitudinal analyses of airway responsiveness and pulmonary function decline. *Am J Respir Crit Care Med* 1996;**154**:5246–9.