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

S Chinn, J P Schouten

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 PD20 or PC20 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 more 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. 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 reproducibility of bronchial responsiveness (BHR) reported the width of the 95% range for a single measurement of an individual’s BHR, expressed as the dose (PD20) or concentration (PC20) estimated to produce a 20% fall in forced expiratory volume in 1 second (FEV1), from under 1 doubling dose or concentration to over 2.5. A 95% range for change is obtained by multiplying the within person standard deviation by .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”.

Limited information on longer term reproducibility has been reported, the width of the 95% range for a single PC20 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 but by epidemiologists as regression dilution, is also a function of the ICC which is often referred to as “reliability”.

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. 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. Hence, despite the ICC being criticised as a measure of agreement, it is clearly of great importance as a measure of repeatability.

Very limited information on ICCs has been given previously. 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 citations32–39 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 PD20 or PC20 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.18

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 PD20 at the first occasion,15 and the smaller recruited participants with wheeze or asthma.16 Three studies achieved low within person variation15 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 PD20 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 PD20 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 PD20 on six occasions, but higher ICCs (0.56 and 0.68) when all first year and all second year pairs were analysed.44 The three long term studies on selected asthmatic subjects had lower within person standard deviations and higher ICCs.10 26 37 The study of aluminium smelter workers included data only for 36 people, those with a 20% fall in FEV1 by the maximum dose of 6.14 µmol at each occasion, in the calculation of ICC for PD20.36 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 FEV1-dose response slope from two studies and long term...
used regression of percentage decline in FEV1 on dose.23 This
PD20-ICC, found a much higher ICC for log dose-response
histamine. The ICCs were 0.89 for slope and 0.81 for PD20.
for each participant who received two or more doses of
data only for persons with two measurable PD20 values in the
The study of aluminium smelter workers, which included
selections. Selection of subjects determines the between subject
larger population studies, had a within person standard
taken greater care over procedures or selected highly
enthusiastic exponents of bronchial challenge may have
doubling doses or concentrations and hence a high ICC. Early
asthmatic patients achieved good repeatability, as indicated
The early studies on short term repeatability in selected
V a r i a t i o n i n I C C
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 concentra-
tion scale as this is the most appropriate for PD20 or PC20,42
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 PD20 has been shown to be high when the dose-response slope is
reciprocally transformed23 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 198643 but, despite this, several later papers reported it, although within person variation was generally also
reported.

Table 2 Short term (less than 4 months) repeatability of PD$_{20}$ or PC$_{20}$ in studies published from 1987 to 1991

<table>
<thead>
<tr>
<th>Year and source</th>
<th>Participants</th>
<th>Age range (years)</th>
<th>Provocation agent and maximum concentration (doubling doses or concentrations)</th>
<th>Time interval</th>
<th>Summary statistic</th>
<th>Within person SD (doubling doses or concentrations)</th>
<th>Between person SD (doubling doses or concentrations)</th>
<th>Intraclass correlation coefficient (ICC)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1987$^*$</td>
<td>90 non-random population sample</td>
<td>18-64</td>
<td>Histamine, 4 µmol</td>
<td>1–7 days</td>
<td>PD$_{20}$-FEV$_1$</td>
<td>0.9</td>
<td>1.9</td>
<td>0.81</td>
</tr>
<tr>
<td>1988$^*$</td>
<td>25 with wheezing or asthma, population sample</td>
<td>18-75</td>
<td>Histamine, 4 µmol</td>
<td>Few weeks</td>
<td>PD$_{20}$-FEV$_1$</td>
<td>1.2</td>
<td>1.3$^*$</td>
<td>0.57$^*$</td>
</tr>
<tr>
<td>1988$^*$</td>
<td>27 with wheezing or asthma, population sample</td>
<td>18-75</td>
<td>Methacholine, 12 µmol</td>
<td>Few weeks</td>
<td>PD$_{20}$-FEV$_1$</td>
<td>1.0</td>
<td>1.9$^*$</td>
<td>0.84$^*$</td>
</tr>
<tr>
<td>1988$^*$</td>
<td>20 asthmatic patients</td>
<td>65-82</td>
<td>Methacholine, 32.65 µmol</td>
<td>Within 10 days</td>
<td>PD$_{20}$-FEV$_1$</td>
<td>0.5</td>
<td>2.7$^*$</td>
<td>0.97$^*$</td>
</tr>
<tr>
<td>1988$^*$</td>
<td>20 asthmatic patients</td>
<td>19-61</td>
<td>Histamine, 10.4 µmol</td>
<td>1–10 days</td>
<td>PD$_{20}$-FEV$_1$</td>
<td>0.3</td>
<td>2.4</td>
<td>0.98</td>
</tr>
<tr>
<td>1989$^*$</td>
<td>20 asthmatic patients</td>
<td>19-66</td>
<td>Methacholine, 24.5 µmol</td>
<td>1–7 days</td>
<td>PD$_{20}$-FEV$_1$</td>
<td>1.0</td>
<td>1.9</td>
<td>0.79</td>
</tr>
<tr>
<td>1989$^*$</td>
<td>19 hospital personnel</td>
<td>23-38</td>
<td>Histamine, 6.83 µmol</td>
<td>1 week</td>
<td>PD$_{20}$-FEV$_1$</td>
<td>0.4</td>
<td>0.4</td>
<td>0.42</td>
</tr>
<tr>
<td>1990$^*$</td>
<td>10 asthmatic patients</td>
<td>18-55</td>
<td>Methacholine, 16 mg/ml</td>
<td>Daily for 5 days</td>
<td>PD$_{20}$-FEV$_1$</td>
<td>1.2</td>
<td>1.6</td>
<td>0.64</td>
</tr>
<tr>
<td>1991$^*$</td>
<td>14 healthy Mean (SD) “responsive”: 22.2 (3.2)</td>
<td>1-14 days</td>
<td>PD$_{20}$-FEV$_1$</td>
<td>0.89</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1991$^*$</td>
<td>20 asthmatic, inexperienced</td>
<td>20-59</td>
<td>Methacholine, 25 µmol</td>
<td>Within 2 weeks</td>
<td>PD$_{20}$-FEV$_1$</td>
<td>0.7</td>
<td>1.9$^*$</td>
<td>0.91$^*$</td>
</tr>
</tbody>
</table>

*Estimated from graphical information. SD, standard deviation.

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

<table>
<thead>
<tr>
<th>Year and source</th>
<th>Participants</th>
<th>Age range (years)</th>
<th>Provocation agent and maximum concentration (doubling doses or concentrations)</th>
<th>Time interval</th>
<th>Summary statistic</th>
<th>Within person SD (doubling doses or concentrations)</th>
<th>Between person SD (doubling doses or concentrations)</th>
<th>Intraclass correlation coefficient (ICC)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1993$^*$</td>
<td>14 asthmatic patients</td>
<td>19-65</td>
<td>Histamine, 6.9 µmol</td>
<td>1 day</td>
<td>PD$_{20}$-FEV$_1$</td>
<td>0.4$^*$</td>
<td>1.8$^*$</td>
<td>0.95</td>
</tr>
<tr>
<td>1993$^*$</td>
<td>20 asthmatic patients</td>
<td>18-65</td>
<td>Methacholine, 32.65 µmol</td>
<td>Within 10 days</td>
<td>PD$_{20}$-FEV$_1$</td>
<td>0.6</td>
<td>2.0$^*$</td>
<td>0.92$^*$</td>
</tr>
<tr>
<td>1995$^*$</td>
<td>19 asthmatic patients</td>
<td>18-45</td>
<td>Methacholine, 32.65 µmol</td>
<td>1 day</td>
<td>log PD$_{20}$-FEV$_1$</td>
<td>0.8$^*$</td>
<td>2.7$^*$</td>
<td>0.92$^*$</td>
</tr>
<tr>
<td>1995$^*$</td>
<td>19 patients</td>
<td>18-57</td>
<td>Histamine, 1.4 µmol</td>
<td>2-7 days</td>
<td>log PD$_{20}$-FEV$_1$</td>
<td>0.6</td>
<td>1.1</td>
<td>0.81</td>
</tr>
<tr>
<td>1996$^*$</td>
<td>11 asthmatic “hyperresponsive” patients</td>
<td>19-62</td>
<td>Methacholine, 35.2 µmol</td>
<td>1-7 days</td>
<td>log PD$_{20}$-FEV$_1$</td>
<td>0.4</td>
<td>1.1$^*$</td>
<td>0.87$^*$</td>
</tr>
<tr>
<td>2001$^*$</td>
<td>34 non-asthmatic</td>
<td>17-78</td>
<td>Methacholine, 64 mg/ml</td>
<td>Within 3 weeks</td>
<td>log PC$_{20}$-FEV$_1$</td>
<td>0.7$^*$</td>
<td>1.4$^*$</td>
<td>0.71$^*$</td>
</tr>
</tbody>
</table>

*Estimated from graphical information. SD, standard deviation.
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 patients21 22 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.23 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 PD20 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 PD20, had a greater ICC than PD20 in one study24 and slightly larger in two others.25 26

Many of the estimates for 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, Sempé gave a 95% confidence interval of 0.70 to 0.96 for an ICC of 0.89 found for ln(PD20) 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 FEV1.14

However, such a high ICC is unlikely to be achieved in larger studies. In the studies which assessed repeatability of FEV1 and BHR in the same subjects, the ICC for BHR was lower than that for FEV111 18 34 40 so that studies of BHR generally require more participants than those on FEV1 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. 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.1 In addition, Schouten and Tager46 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 4 Long term (more than 4 months) repeatability of BHR

<table>
<thead>
<tr>
<th>Year and source</th>
<th>Participants</th>
<th>Age range (years)</th>
<th>Provocation agent and maximum concentration/ cumulative dose</th>
<th>Time interval</th>
<th>Summary statistic</th>
<th>Within person SD (doubling doses or concentrations)</th>
<th>Between person SD (doubling doses or concentrations)</th>
<th>Intraclass correlation coefficient (ICC)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1989*</td>
<td>35 asthmatic</td>
<td>Not stated</td>
<td>Histamine, 16 mg/ml</td>
<td>10-28 months</td>
<td>PC20/FEV1</td>
<td>1.0*</td>
<td>1.7*</td>
<td>0.74*</td>
</tr>
<tr>
<td>1990*</td>
<td>10 atopic asthmatic patients</td>
<td>23-41 at baseline</td>
<td>Histamine, 8.0 mg/ml</td>
<td>14-28 months</td>
<td>PC20/FEV1</td>
<td>1.6*</td>
<td>1.4*</td>
<td>0.84</td>
</tr>
<tr>
<td>1991*</td>
<td>38 “responsive” workers</td>
<td>Not stated</td>
<td>Methacholine, 6.14 μmol</td>
<td>1 year</td>
<td>PD20/FEV1</td>
<td>0.28</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1993*</td>
<td>1413 asymptomatic population sample</td>
<td>15-54 at baseline</td>
<td>Histamine, 32 mg/ml</td>
<td>3-22 years</td>
<td>PC10/FEV1</td>
<td>0.8</td>
<td></td>
<td>0.63</td>
</tr>
<tr>
<td>1993*</td>
<td>803 symptomatic population sample</td>
<td>15-54 at baseline</td>
<td>Histamine, 32 mg/ml</td>
<td>3-22 years</td>
<td>PC10/FEV1</td>
<td>0.9</td>
<td></td>
<td>0.77</td>
</tr>
<tr>
<td>1994*</td>
<td>1273 population sample</td>
<td>15-54 at baseline</td>
<td>Histamine, 32 mg/ml</td>
<td>3-22 years</td>
<td>PC10/FEV1</td>
<td>1.0</td>
<td>0.8</td>
<td>0.63</td>
</tr>
<tr>
<td>1994*</td>
<td>1273 population sample</td>
<td>18-75</td>
<td>Methacholine, 247 μmol</td>
<td>4-24 months</td>
<td>PC10/FEV1</td>
<td>1.1</td>
<td>1.1, calculated from ICC and within person SD</td>
<td>0.75</td>
</tr>
<tr>
<td>1996*</td>
<td>10 asthmatic, PC20 &lt;9 mg/ml</td>
<td>25-82</td>
<td>Histamine, 16 mg/ml</td>
<td>6 months</td>
<td>PC20/FEV1</td>
<td>0.7*</td>
<td>1.4*</td>
<td>0.80*</td>
</tr>
<tr>
<td>1997*</td>
<td>88 healthy working adults</td>
<td>Not stated</td>
<td>Methacholine, 13.0 μmol</td>
<td>1-3 years</td>
<td>PC20/FEV1</td>
<td>1.8</td>
<td>1.6</td>
<td>0.65</td>
</tr>
</tbody>
</table>

*Estimated from graphical information.

SD, standard deviation.

V/V unpublished results from Vlagtwedden/Vlaardingen study; published results also included.

### Table 5 Intraclass correlation coefficients for measures of FEV1-dose-response slope

<table>
<thead>
<tr>
<th>Year and source</th>
<th>Participants</th>
<th>Age range (years)</th>
<th>Provocation agent and maximum concentration/ cumulative dose</th>
<th>Time interval</th>
<th>Summary statistic</th>
<th>Intraclass correlation coefficient (ICC)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1990*</td>
<td>726 workers</td>
<td>Not stated</td>
<td>Methacholine, 6.14 μmol</td>
<td>1 year</td>
<td>Log dose-response slope</td>
<td>0.73</td>
</tr>
<tr>
<td>1991*</td>
<td>14 healthy “responsive”</td>
<td>Mean (SD) 22.2 [3.2]</td>
<td>Methacholine, 8.5 μmol</td>
<td>1-14 days</td>
<td>Log dose-response slope</td>
<td>0.79</td>
</tr>
<tr>
<td>1991*</td>
<td>16 healthy “non-responsive”</td>
<td>Mean (SD) 25.4 [3.3]</td>
<td>Methacholine, 8.5 μmol</td>
<td>1-14 days</td>
<td>Log dose-response slope</td>
<td>0.80</td>
</tr>
<tr>
<td>1993*</td>
<td>104 population sample</td>
<td>18-64</td>
<td>Methacholine, 4 μmol</td>
<td>1-14 days</td>
<td>Log dose-response slope</td>
<td>0.80</td>
</tr>
<tr>
<td>1994*</td>
<td>67 general practice on 6 occasions</td>
<td>18-75</td>
<td>Methacholine, 247 μmol</td>
<td>4-24 months</td>
<td>Log dose-response slope</td>
<td>0.75</td>
</tr>
<tr>
<td>1994*</td>
<td>12 healthy smokers</td>
<td>27.4 [3.4]</td>
<td>Histamine, 16 mg/ml</td>
<td>1-3 days</td>
<td>Dose-response slope</td>
<td>0.97</td>
</tr>
<tr>
<td>1997*</td>
<td>88 healthy working adults</td>
<td>Mean (SD) 27.4 [3.4]</td>
<td>Methacholine, 13.06 μmol in repeatability data</td>
<td>1-3 years</td>
<td>Log dose-response slope</td>
<td>0.54</td>
</tr>
</tbody>
</table>

FEV1, 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 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.  

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 FEV1, 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 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 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.

**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. PD20 or PC20 are known only to be above the maximum dose or concentration (that is, “censored”) when a 20% fall in FEV1 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. This is reinforced by ICCs for the dose-response slope being at least as high and probably greater than those for PD20. 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 FEV1 as the short term ICC for FEV1 can be presumed to be over 0.917 0.918 0.919 0.920 and has been reported to be 0.89 over 1–3 years. 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 FEV1. 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 FEV1 is a cause or a result of increased airway responsiveness and stated that longitudinal analyses are necessary to answer the question. We can add to this that either multiple measurements of BHR should be made to increase precision 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**


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

Thorax 2005 60: 395-400
doi: 10.1136/thx.2004.039230

Updated information and services can be found at:
http://thorax.bmj.com/content/60/5/395

These include:

References
This article cites 47 articles, 19 of which you can access for free at:
http://thorax.bmj.com/content/60/5/395#BIBL

Email alerting service
Receive free email alerts when new articles cite this article. Sign up in the box at the top right corner of the online article.

Topic Collections
Articles on similar topics can be found in the following collections

- Airway biology (1100)
- Asthma (1782)
- Epidemiologic studies (1829)
- Internet (104)
- Lung function (773)

Notes

To request permissions go to:
http://group.bmj.com/group/rights-licensing/permissions

To order reprints go to:
http://journals.bmj.com/cgi/reprintform

To subscribe to BMJ go to:
http://group.bmj.com/subscribe/