Eur Respir J 2000; 16: 472–475
Printed in UK – all rights reserved

European Respiratory Journal
ISSN 0903-1936

Observer variability of lung function measurements in 2–6-yr-old children

B. Klug, K.G. Nielsen, H. Bisgaard

Observer variability of lung function measurements in 2–6-yr-old children. B. Klug, K.G. Nielsen, H. Bisgaard. ©Eur Respir J Ltd 2000.
ABSTRACT: The aim of this study was to assess the within-observer and between-observer variability of lung function measurements in children aged 2–6 yrs.

Two observers examined 22 asthmatic children independently according to a predefined protocol. Each observer obtained duplicate measurements of respiratory resistance by the interrupter technique (R_{int}), respiratory resistance ($R_{\mathrm{rs,5}}$) and reactance ($X_{\mathrm{rs,5}}$) at 5 Hz by the impulse oscillation technique and the specific airway resistance (sR_{aw}) by whole body plethysmography.

The within-subject SD (SDw) was not significantly different in the two observers. The ratio SDw between observers/mean SDw within observers was 0.94, 1.25, 1.35 and 2.86 for $X_{\mathrm{rs,5}}$, $R_{\mathrm{rs,5}}$, sR_{aw} and R_{int}, respectively, indicating greater between-observer variability of the latter. The systematic difference between observers assessed by the difference between observer means (expressed as a percentage of their mean value) was 11, 7, 6 and 2% for $X_{\mathrm{rs,5}}$, sR_{aw}, $R_{\mathrm{rs,5}}$ and R_{int}, respectively. These differences were statistically significant, except that for R_{int}.

In conclusion, specific airway resistance, impulse oscillation technique and respiratory resistance assessed by the interrupter technique measurements in young children are subject to influence by the observer, and the random variability between observers appears to be particularly great for respiratory resistance assessed by the interrupter technique. The authors suggest that the between-observer variability should be investigated when evaluating novel methods for testing lung function.
Eur Respir J 2000; 16: 472–475.

Dept of Paediatrics 5003, National University Hospital, Rigshospitalet, Copenhagen, Denmark.

Correspondence: B. Klug
Dept of Paediatrics
Hvidovre University Hospital
Kettegaard Alle 30
Copenhagen 2650
Denmark.
Fax: 45 36323770

Keywords: Preschool children
reproducibility of results
respiratory function tests

Received: December 22 1999
Accepted after revision May 30 2000

Measurement of lung function has recently been successfully applied in unsedated young children; lung function testing in young children is, therefore, likely to become more widely employed. Respiratory resistance assessed by the interrupter technique (R_{int}), the impulse oscillation technique (IOS) and a modified method for measuring specific airway resistance (sR_{aw}) by whole body plethysmography have been previously evaluated in young children [1–3]. It was found that measurements can be achieved without prior training in the majority of young children, down to the age of 2 yrs, and that the reproducibility of the measurements is satisfactory and independent of age.

The sources of variability and their relative magnitude must be considered when interpreting the outcome of measurements. Information about variability is often limited to data on within-observer variability, which assesses the reproducibility of measurements obtained by a single observer in the same patient. Despite a standardized protocol, observers may execute the measurements in a slightly different manner and the interpretation of the outcome of the measurements may also differ. Such differences will be disclosed by assessing the between-observer variability by examining the agreement between measurements made by different observers in the same patient. Few data are available on the influence of the observer on the outcome of lung function measurements [4], whereas a number of studies have investigated the agreement between observers reassessing lung function data already recorded [5, 6].

Standardization of procedures and employment of criteria for the acceptance and rejection of data serve to reduce variability, which ideally should be similar within and between observers. The purpose of this study was to compare the variability within and between observers in lung function measurements in young children, applying R_{int}, IOS and sR_{aw} measurements.

Subjects and methods

Subjects

The children were recruited from the paediatric asthma outpatient clinic at Rigshospitalet. Children aged 2–6 yrs who had experienced three or more episodes of wheezing after the age of 1 yr were eligible for the study. The study was approved by the local medical ethics committee, and written informed consent was obtained from the parents of the children.

Methods

Measurements of lung function were obtained by two observers (A: Klug B.K. and B: Nielson K.G.N.) with extensive experience in testing lung function in young children. Each observer obtained duplicate measurements in a fixed sequence: R_{int}, IOS and sR_{aw}. The order of the two observers was determined randomly and they were blinded with regard to each other's measurements. A protocol describing the procedures and the criteria for acceptance or rejection of measurements was available for each method. Measurements should be carried out during tidal breathing and continue until a predefined amount of data is collected from consecutive breaths, free of artefacts, as previously described in detail [1, 2]. Specific criteria were employed for each method: R_{int} was calculated as the mean of five single measurements obtained at every second inspiratory phase of ten consecutive breaths. Each measurement should be within ±15% of the mean value of the five. The IOS primary data were recorded during 30 s of uninterrupted breathing at a respiratory frequency within the range 20–40 breaths·min^{-1}. The results were calculated from the primary data immediately after completing data acquisition. Plethysmographic measurements of sR_{aw} were carried out at a respiratory frequency within the range 30–45 breaths·min^{-1}. sR_{aw} was measured from the relation between simultaneous variations in respiratory flow and in plethysmographic volume, omitting measurement of thoracic gas volume [7, 8]. sR_{aw} was calculated as the median value derived from five consecutive specific resistance loops which, as judged by the slope and shape, had similar appearances as on the on-line monitor. From these five loops, three different estimates of sR_{aw} were obtained using the line connecting: 1) the points of maximum change in plethysmographic volume during inspiration and expiration (s$R_{\mathrm{aw},\Delta V,\max}$); 2) the intercept of the specific resistance loop at inspiratory and expiratory flow rates of 0.2 L·s^{-1} (s$R_{\mathrm{aw},0.2}$); and 3) the points at which flow was 50% of the maximum during inspiration and expiration (s$R_{\mathrm{aw},50\%}$) [3]. The observers employed this protocol during data acquisition and subsequently no editing or exclusion of data was allowed.

Analysis of data

The reproducibility of the measurements was estimated using the within-subject SD (SDw). For each observer, the within-observer SDw was calculated as the SD of the differences between the two measurements obtained in all subjects divided by $\sqrt{2}$. The between-observer SDw was calculated as the SD of the differences between the means of the two measurements obtained by each observer in all subjects divided by $\sqrt{2}$. The difference between paired measurements were plotted against their mean in order to examine whether the variability was independent of the magnitude of the measurements [9], and 95% limits of agreement were calculated as suggested by CHINN [10]. Differences between observer means and between the first and second sequence of measurements were analysed by analysis of variance. The number of patients required, in a cross-over study, to detect a difference equal to 1 SDw with a power of 80% and a significance level of 5% was calculated for the individual lung function indices, depending on the observer and the combination of observers. The SDw used in these calculations were those established in a recent study in healthy children [3]. The F-test was used for comparing reproducibility within and between observers. A p-value of <0.05 was considered statistically significant.

Table 1. – Variability of lung function measurements within observers (A and B) and between observers (A-B) in 22 children

Method	Observer	Group mean±SD	ΔWithin subject* mean±SD	95% limits of agreement	SDw
s$R_{\mathrm{aw},\Delta V_{\max}}$ kPa·s	A	1.41±0.39	0.05±0.12	-0.19–0.29	0.087
	B	1.32±0.31	-0.04±0.18	-0.42–0.34	0.126
	A-B		0.09±0.19 0.02–0.15	-0.49–0.31	0.144
s$R_{\mathrm{aw},0.2}$ kPa·L^{-1}·s	A	0.84±0.27	0.02±0.11	-0.20–0.25	0.076
	B	0.86±0.27	-0.03±0.11	-0.25–0.20	0.063
	A-B		-0.02±0.12 -0.06–0.03	-0.23–0.27	0.082
s$R_{\mathrm{aw},50\%}$ kPa·L^{-1}·s	A	0.93±0.29	0.01±0.09	-0.18–0.19	0.061
	B	0.89±0.28	-0.02±0.11	-0.25–0.21	0.076
	A-B		0.05±0.12 0.01–0.09	-0.29–0.20	0.086
R_{int} kPa·L^{-1}·s	A	1.27±0.22	-0.03±0.12	-0.28–0.22	0.085
	B	1.25±0.29	0.00±0.09	-0.19–0.19	0.063
	A-B		0.02±0.31 -0.06–0.09	-0.66–0.61	0.212
$R_{\mathrm{rs},5}$ kPa·L^{-1}·s	A	1.21±0.23	-0.01±0.13	-0.28–0.37	0.090
	B	1.14±0.22	0.04±0.16	-0.29–0.37	0.113
	A-B		0.07±0.17 0.02–0.12	-0.43–0.28	0.127
$X_{\mathrm{rs},5}$ kPa·L^{-1}·s	A	-0.37±0.12	-0.04±0.13	-0.30–0.23	0.091
	B	-0.33±0.14	0.00±0.10	-0.20–0.20	0.066
	A-B		-0.04±0.10 -0.08–-0.01	-0.16–0.25	0.074

Values in parentheses represent 95% confidence intervals. *: calculated as the mean difference between two measurements obtained by each observer for A and B, and as the difference between the mean value of duplicated measurements obtained by each oberver for A and B. Δ: difference; SDw: within-subject SD; s$R_{\mathrm{aw},\Delta V_{\max}}$: specific airway resistance (sR_{aw}) estimated from maximum change in plethysmographic volume during inspiration and expiration; s$R_{\mathrm{aw},0.2}$: sR_{aw} estimated at inspiratory and expiratory flow rates at 0.2 L·s^{-1}; s$R_{\mathrm{aw},50\%}$: sR_{aw} estimated at 50% maximum flow during inspiration and expiration; R_{int}: respiratory resistance measured by the interrupter technique; $R_{\mathrm{rs},5}$: respiratory resistance at 5 Hz; $X_{\mathrm{rs},5}$: respiratory reactance at 5 Hz. Specific airway resistance (sR_{aw}) is calculated from the specific resistance loop using the inspiratory and expiratory flow points at which the change of plethysmographic volume is maximal ($\Delta V_{\max}$).

Results

Twenty-five children were enrolled. Three children did not complete any measurements with one of the observers. Twenty-two children, 11 males and 11 females, with a mean age of 4.8 yrs (range 2.6–6.8 yrs) completed the study. Of the 22 children who completed the measurements, six were familiar with the measurements and 16 were naive to the measurements. Observer A and B commenced the first sequence of measurements in 12 and 10 children, respectively. Measurements by whole body plethysmography were not available for the three children who failed to complete the measurements. For sR_{aw} measurements performed by observer A and B the mean±SD respiratory frequency was 41±3.8 and 39±4.4 breaths·min^{-1}, and, for IOS measurements, 27±5.5 and 28±5.4 breaths·min^{-1}, respectively. The within-observer and between observer variability was independent of the level of the measurements. The SD$_w$ of the measurements made by the second observer were greater but not significantly different from those made by the first observer.

A significant difference between observers was found for the mean $sR_{aw,\Delta V_{max}}$, $sR_{aw,50\%}$, $X_{rs,5}$ and $R_{rs,5}$, but not for R_{int} and $sR_{aw,0.2}$ (table 1). This difference between observers expressed as a percentage of the observer means was 11, 7, 6, 5, 2 and 2% for $X_{rs,5}$, $sR_{aw,\Delta V_{max}}$, $R_{rs,5}$, $sR_{aw,50\%}$, $sR_{aw,0.2}$ and R_{int}, respectively.

The within-observer SD$_w$ was not significantly different in observers A and B. In order to compare the magnitude of the between-observer variability of the different lung function methods, the ratio of the SD$_w$ between and within observers (SD$_w$ between observers/mean SD$_w$ within observers) was calculated. This ratio is independent of the absolute values of the measurements obtained by the different methods. If the variability were purely random, without influence from different observers, the ratio would be ~1. The more the between-observer variability exceeds the within-observer variability, the more the ratio will exceed 1. The ratio was: 0.94, 1.18, 1.25, 1.26, 1.35 and 2.86 for $X_{rs,5}$, $sR_{aw,0.2}$, $R_{rs,5}$, $sR_{aw,50\%}$, $sR_{aw,\Delta V_{max}}$, and R_{int}, respectively, indicating a greater influence of between-observer variability on the latter method. The 95% limits of agreement within and between observers are given in table 1.

Table 2. – Number of patients required for different methods in a cross-over study, dependent on the observer(s)*

	One observer		Two observers	
	A	B	AB$^+$	A/B$^{\#}$
$sR_{aw,\Delta V,max}$	16	26	21	39
$sR_{aw,0.2}$	12	12	12	15
$sR_{aw,50\%}$	8	11	9	16
R_{int}	20	28	24	134
$R_{rs,5}$	13	27	20	35
$X_{rs,5}$	9	10	10	10

*: to detect a difference equal to 1 within-subject SD for each method, with a power of 80% and a significance level of 5%; $^+$: each observer examines the same 50% of patients throughout the study; $^{\#}$: observer A examines all patients before cross-over and observer B examines all patients after cross-over. $sR_{aw,\Delta V_{max}}$: specific airway resistance (sR_{aw}) estimated from maximum change in plethysmographic volume during inspiration and expiration; $sR_{aw,0.2}$: sR_{aw} estimated at inspiratory and expiratory flow rates at 0.2 L·s^{-1}; $sR_{aw,50\%}$: sR_{aw} estimated at 50% maximum flow during inspiration and expiration; R_{int}: respiratory resistance measured by the interrupter technique; $R_{rs,5}$: respiratory resistance at 5 Hz; $X_{rs,5}$: respiratory reactance at 5 Hz.

The consequence of the differences in variability within and between observers is illustrated by the number of patients required to detect a difference of 1 SD$_w$ in a cross-over study, dependent on the observer(s). Table 2 shows that, were the two observers to perform the measurements in a nonrandom manner, the number of patients would increase, and, in the case of R_{int} this increase would be substantial.

Discussion

The present results suggest that sR_{aw}, R_{int} and IOS measurements in young children are subject to influence by the observer, despite employment of a protocol standardizing the measurement procedures and the criteria for acceptance or rejection of data. It was found that within-subject variability was comparable for the two observers but greater between observers than with observers for all methods, except for $X_{rs,5}$.

The between-observer variability appeared not to be purely random, as indicated by the finding of a statistically significant difference between the mean $R_{rs,5}$, $X_{rs,5}$, $sR_{aw,\Delta V_{max}}$ and $sR_{aw,50\%}$ obtained by the two observers, but none between the mean R_{int} and $sR_{aw,0.2}$. This observation suggests that the between-observer variability of sR_{aw} may be reduced by calculating sR_{aw} from the linear part of the specific resistance loop.

The significant difference between the mean observer values found for most of the lung function indices indicate a systematic difference between the observers. The explanation for this difference is unclear, but the practical execution of the measurements may have differed between the observers. sR_{aw} and IOS measurements are influenced by respiratory frequency; however, the respiratory frequency of the children did not show any systematic difference between the observer's. Bias within each observers paired measurements may have contributed to within-subject variability being greater between observers than within observers. In many children, measurements fluctuate immediately after commencing the lung function test, and, in some cases, may continue to vary over a longer period of time. Therefore, the observer has to monitor the traces on the on-line display for a period of time prior to the collection of data in order to define the value that appears to be representative. This decision may be subject to bias since it cannot be derived unambiguously from a predefined protocol. Hence, knowledge about the outcome of the first measurement might bias the observer, during the second measurement such that data showing good agreement with the first measurement, are more readily accepted and apparently deviating data are not retained.

Difficulties in establishing a representative measurement of lung function are common to methods that measure lung function during tidal breathing. By contrast, when measuring lung function by forced expiratory flows, the representative value is defined as the highest value selected from a number of attempts and the risk of overestimating lung function is minimal [11], although underestimation may occur if the expiratory effort is submaximal [12]. Measurements of airway/respiratory resistance by tidal breathing methods are influenced by factors that may vary within short time intervals, *e.g.* respiratory frequency, airflow, lung volume and glottis opening.

The magnitude of between-observer variability has practical implications when defining what may be considered a significant change in lung function, *e.g.* in response to antiasthmatic medication. The within-subject variability of R_{int} showed no systematic difference between observers but R_{int} was markedly greater between observers than within observers, as reflected by a substantial widening of the 95% limits of agreement. The consequence of the high between-observer variability of R_{int} is illustrated by the fact that the number of patients needed to demonstrate a predefined change in lung function increases substantially if the observers are not randomly assigned to perform the lung function tests (table 2).

The present results indicate that specific airway resistance; respiratory resistance assessed by the interrupter technique and impulse oscillation technique measurements in young children are influenced by the observer. Further studies are needed to investigate how to improve the present protocol in order to reduce the variability between observers. The authors suggest that assessment of between-observer variability should be an integratal part of the investigation of novel lung function testing methods.

References

1. Bisgaard H, Klug B. Lung function measurement in awake young children. *Eur Respir J* 1995; 8: 2067–2075.
2. Klug B, Bisgaard H. Measurement of lung function in awake 2–4 year-old asthmatic children during methacholine challenge and acute asthma: a comparison of the impulse oscillation technique, the interrupter technique, and transcutaneous measurement of oxygen *versus* whole-body plethysmography. *Pediatr Pulmonol* 1996; 21: 290–300.
3. Klug B, Bisgaard H. Specific airway resistance, interrupter resistance, and respiratory impedance in healthy children aged 2–7 years. *Pediatr Pulmonol* 1998; 25: 322–331.
4. Groth S, Dirksen A, Dirksen H, Rossing N. Intraindividual variation and effect of learning in lung function examinations. A population study. *Bull Eur Physiopathol Respir* 1986; 22: 35–42.
5. Teculescu DB, Bohadana AB, Peslin R, Pino J, Jansen da Silva JM. Variability, reproducibility and observer difference of body plethysmographic measurements. *Clin Physiol* 1982; 2: 127–138.
6. Dundas I, Beardsmore C, Wellman T, Stocks J. A collaborative study of infant respiratory function testing. *Eur Respir J* 1998; 12: 944–953.
7. Dab I, Alexander F. A simplified approach to the measurement of specific airway resistance. *Pediatr Res* 1976; 10: 998–999.
8. Klug B, Bisgaard H. Measurement of the specific airway resistance by plethysmography in young children accompanied by an adult. *Eur Respir J* 1997; 10: 1599–1605.
9. Bland JM, Altman DG. Statistical methods for assessing agreement between two methods of clinical measurement. *Lancet* 1986; 1: 307–310.
10. Chinn S. Statistics in respiratory medicine. 2. Repeatability and method comparison. *Thorax* 1991; 46: 454–456.
11. Quanjer PH, Tammeling GJ, Cotes JE, Pedersen OF, Peslin R, Yernault JC. Lung volumes and forced ventilatory flows. Report Working Party Standardization of Lung Function Tests, European Community for Steel and Coal. Official Statement of the European Respiratory Society. *Eur Respir J* 1993; 6: Suppl. 16: 5–40.
12. Kanengiser S, Dozor AJ. Forced expiratory maneuvers in children aged 3 to 5 years. *Pediatr Pulmonol* 1994; 18: 144–149.