# Controlled *versus* free breathing for multiple breath nitrogen washout in healthy adults

Blake M. Handley[1,2], Edward Jeagal[2], Robin E. Schoeffel[1], Tanya Badal[2], David G. Chapman [2,3], Catherine E. Farrow[2,4], Gregory G. King[1,2,5], Paul D. Robinson [6], Stephen Milne [5,7,8] and Cindy Thamrin [2,5,8]

**Affiliations**: [1]Dept of Respiratory Medicine, Royal North Shore Hospital, St Leonards, NSW, Australia. [2]Woolcock Institute of Medical Research, University of Sydney, Glebe, NSW, Australia. [3]Translational Airways Group, School of Life Sciences, Faculty of Science, University of Technology Sydney, Ultimo, NSW, Australia. [4]Dept of Respiratory and Sleep Medicine, Westmead Hospital, Westmead, NSW, Australia. [5]Faculty of Medicine and Health, University of Sydney, Camperdown, NSW, Australia. [6]Dept of Respiratory Medicine, The Children's Hospital at Westmead, Westmead, NSW, Australia. [7]Centre for Heart Lung Innovation, St Paul's Hospital and Division of Respiratory Medicine, University of British Columbia, Vancouver, BC, Canada. [8]These authors contributed equally.

**Correspondence**: Cindy Thamrin, Airway Physiology and Imaging Group, Woolcock Institute of Medical Research, 431 Glebe Point Rd, Glebe NSW 2037, Australia. E-mail: cindy.thamrin@sydney.edu.au

ABSTRACT Multiple breath nitrogen washout (MBNW) quantifies ventilation heterogeneity. Two distinct protocols are currently used for MBNW testing: "controlled breathing", with targeted tidal volume ($V_T$) and respiratory rate (RR); and "free breathing", with no constraints on breathing pattern. Indices derived from the two protocols (functional residual capacity (FRC), lung clearance index (LCI), $S_{cond}$, $S_{acin}$) have not been directly compared in adults. We aimed to determine whether MBNW indices are comparable between protocols, to identify factors underlying any between-protocol differences and to determine the between-session variabilities of each protocol.

We performed MBNW testing by both protocols in 27 healthy adult volunteers, applying the currently proposed correction for $V_T$ to $S_{cond}$ and $S_{acin}$ derived from free breathing. To establish between-session variability, we repeated testing in 15 volunteers within 3 months.

While FRC was comparable between controlled *versus* free breathing (3.17 (0.98) *versus* 3.18 (0.94) L, p=0.88), indices of ventilation heterogeneity derived from the two protocols were not, with poor correlation for $S_{cond}$ (r=0.18, p=0.36) and significant bias for $S_{acin}$ (0.057 (0.021) $L^{-1}$ *versus* 0.085 (0.038) $L^{-1}$, p=0.0004). Between-protocol differences in $S_{acin}$ were related to differences in the breathing pattern, *i.e.* $V_T$ (p=0.004) and RR (p=0.01), rather than FRC. FRC and LCI showed good between-session repeatability, but $S_{cond}$ and $S_{acin}$ from free breathing showed poor repeatability with wide limits of agreement.

These findings have implications for the ongoing clinical implementation of MBNW, as they demonstrate that $S_{cond}$ and $S_{acin}$ from free breathing, despite $V_T$ correction, are not equivalent to the controlled breathing protocol. The poor between-session repeatability of $S_{cond}$ during free breathing may limit its clinical utility.

**@ERSpublications**
**Phase 3 slopes indices derived from "free breathing" and "controlled breathing" MBNW protocols are not comparable, and differences are related to breathing patterns. These findings have implications for the ongoing clinical implementation of MBNW.** https://bit.ly/35oQYnW

**Cite this article as:** Handley BM, Jeagal E, Schoeffel RE, *et al*. Controlled *versus* free breathing for multiple breath nitrogen washout in healthy adults. *ERJ Open Res* 2021; 7: 00435-2020 [https://doi.org/10.1183/23120541.00435-2020].

## Introduction

Increased ventilation heterogeneity is a characteristic physiological abnormality in respiratory diseases such as asthma and COPD [1, 2]. The multiple breath nitrogen washout (MBNW) test is an increasingly available method of quantifying ventilation heterogeneity. MBNW is conducted by breathing 100% oxygen ($O_2$) which then "washes out" the resident nitrogen ($N_2$) in the lung. Analysis of the exhaled $N_2$ concentration *versus* exhaled volume of each breath then allows calculation of a number of parameters: a global measure of heterogeneity known as the lung clearance index (LCI); heterogeneity arising predominantly within the conducting airways ($S_{cond}$); heterogeneity arising predominantly in the more distal/intra-acinar airways ($S_{acin}$); and functional residual capacity (FRC).

Two distinct breathing protocols are currently used for MBNW testing. One is a "controlled breathing" protocol whereby the patient maintains a consistent tidal volume ($V_T$) (∼1 L) and respiratory rate (RR) (8–12 breaths·min$^{-1}$) [3] through the use of a visual incentive screen and real-time feedback from the test operator. These breathing constraints are applied to all individuals equally, regardless of lung size; consequently, a given patient's breathing pattern during the test may deviate significantly from their usual resting breathing pattern. The alternative is a "free breathing" protocol, with no constraints on $V_T$ or RR, *i.e.* the individual determines their own breathing depth and frequency. While this makes the method more suited to paediatric testing [3], variability in breathing patterns within or between individuals may have an impact on indices of ventilation heterogeneity [4–6] and their repeatability. A correction for lung size and breath-to-breath variation in $V_T$ has been proposed for this method [7], but this has not been formally validated against a controlled breathing approach. Standard reference equations are available for both methods to guide interpretation of MBNW indices, which account for factors such as age, height and sex [8–11], but little is known about their comparability.

Differences in how controlled and free breathing MBNW protocols are conducted and analysed may therefore limit the comparability of indices derived from them. A direct comparison of the protocols in children [12] showed significant differences in LCI and $S_{cond}$; the authors concluded that this was due to a reduction in end-expiratory lung volume (EELV) induced by controlled breathing. However, the results suggested that other patient-related factors (*e.g.* lung size, body weight) may contribute to the discrepancies. Indices derived from the two protocols have not been directly compared in adults, nor have their respective between-session variabilities.

Therefore, in a sample of healthy adults, we aimed to determine: 1) whether controlled and free breathing MBNW protocols provide equivalent FRC and indices of ventilation heterogeneity; 2) the influence of patient-related factors (anthropometrics and/or breathing pattern) on any observable differences; and 3) if the choice of protocol influences between-session variability, which would allow us to interpret any between-protocol differences. Our overall hypothesis was that controlled and free breathing MBNW would produce comparable measurements of ventilation heterogeneity, with the secondary hypothesis that any differences would be, in part, due to patient-related factors.

## Methods

### Study overview

We recruited and studied volunteers aged ⩾18 years in the respiratory function laboratories at two sites (Royal North Shore Hospital (RNSH) and Woolcock Institute of Medical Research, Sydney, Australia; see supplementary material for inclusion/exclusion criteria) from a convenience sample of volunteers (predominantly hospital staff) between April 2018 and November 2019. Our target sample size was n=25, comparable to the control group of a previous study [12]. The RNSH Human Research Ethics Committee approved the study (LNR/16/HAWKE/11). After obtaining written informed consent to be tested and for their data to be used for research purposes, participants first underwent standard (pre-bronchodilator) lung function testing to confirm they had normal lung function. After a period of rest of at least 10 min, participants underwent MBNW testing by both breathing protocols conducted in a randomised order. Each participant was invited to return for repeat testing, at their convenience, within the next 3 months. This follow-up testing was performed in an identical manner, including the order of MBNW protocol testing, for consistency.

### Standard lung function testing

We measured spirometry and lung volumes according to American Thoracic Society/European Respiratory Society (ATS/ERS) quality criteria using a Jaeger MasterScreen PFT (Vyaire Medical GmbH, Hoechberg, Germany) device, with comparison to reference values [13, 14] for plethysmography.

### MBNW testing

Detailed descriptions of the device and testing procedure are found in the supplementary material. After at least 10 min of rest for the participant, we conducted MBNW tests using the Exhalyzer D with Spiroware v3.1.6 (Eco Medics AG, Duernten, Switzerland).

Participants performed both controlled and free breathing protocols in succession, in a random order (as determined by a computer-based random number generator), according to current international consensus recommendations [3]. All trials started with a period of normal relaxed breathing on room air in order to establish a stable EELV. For the controlled breathing protocol, the operator then instructed the participant to breathe at a $V_T$ of 0.95–1.3 L and at RR 8–12/min with the use of the visual incentive screen within Spiroware (described by VERBANCK et al. [15]); once the operator was satisfied that a stable breathing pattern and EELV had been achieved, they commenced the washout by switching the circuit to 100% $O_2$, and the participant maintained the same breathing pattern for the duration of the washout phase. For the free breathing protocol, the operator instructed the participant to "continue to breathe in a normal relaxed manner" through both the pre-phase and washout phase, with the visual incentive screen switched off and without any additional coaching during the trial. Participants repeated the test until at least three technically acceptable trials with FRC values within ±10% of the mean were achieved for each protocol.

### MBNW analysis

Full details are provided in the supplementary material. We analysed MBNW data using Spiroware software (v.3.1.6). While the test operator performed preliminary analysis on individual trials during the testing session, a single investigator re-analysed the data *post hoc* for all participants as a batch in order to ensure a consistent approach to analysis.

To differentiate between measurements made by the different protocols, we have subscripted all indices with CB or FB to indicate controlled and free breathing, respectively. Under the free breathing protocol, $S_{cond}$ and $S_{acin}$ are adjusted for $V_T$ and are thus denoted in the literature and Spiroware software as $S_{cond}*V_T$ and $S_{acin}*V_T$; however, we refer to them here simply as $S_{cond_{FB}}$ and $S_{acin_{FB}}$, respectively.

### Statistical analyses

We compared FRC, LCI, $S_{cond}$ and $S_{acin}$ measured by the two breathing protocols using paired *t*-tests and Pearson correlation. Additionally, we compared FRC from either protocol against the gold-standard $FRC_{pleth}$. To investigate for bias, we generated Bland–Altman plots as the between-protocol difference (free breathing minus controlled breathing) *versus* the average, plotting the mean difference and 95% limits of agreement (95% limits of agreement (LOA)). We then performed linear regression of the difference *versus* average to determine any proportional bias. To examine the effects of various predictors (age, sex, height, BMI, mean RR from free breathing, mean $V_T$ from free breathing and $FRC_{pleth}$) on between-protocol difference, we performed linear regression of the difference *versus* each predictor. To determine within-session variability, we calculated the coefficient of variation (CoV) from the three washout trials. To determine between-session variability, we calculated the difference (Visit 2 minus Visit 1) and 95% LOA separately for each protocol. We also report the between-session intra-class correlation coefficients (ICC), calculated using a two-way mixed effects ANOVA model based on absolute agreement, multiple measurements (k=3) [16]. We set statistical significance at p<0.05. Results are presented as mean (standard deviation (SD)) unless otherwise stated.

## Results

### Participant characteristics and breathing patterns

We studied 27 non-smoking, healthy volunteers (22 at the RNSH site, 5 at the WIMR site) with a median age of 34 (range 19–65) years and spirometry/lung volumes within the limits of normal (table 1). Since MBNW parameters are likely to be device- and protocol-specific [17, 18], we were unable to compare all measurements to a single set of reference equations. However, MBNW indices from the free breathing protocol were within the ranges of normal derived from the same testing device (supplementary table S1) [9].

Compared to controlled breathing, free breathing produced smaller mean $V_T$ (mean (SD)) difference −0.24 (0.33) L, p=0.0006), faster mean RR (1.6 (3.4) breaths·min$^{-1}$, p=0.02), smaller minute ventilation (−2.4 (3.1) L·min$^{-1}$, p<0.0001) and greater cumulative expired volume (CEV) (1.84 L, p=0.006), but similar washout times (p=0.17). Four out of the 27 participants required coaching prior to commencing the free breathing protocol due to inadequate phase III. The average proportions of washout breaths excluded from analysis due to inadequate phase III were 2.7% for controlled breathing and 3.9% for free breathing.

### FRC was comparable between protocols and with plethysmography

There was no significant difference in mean FRC measured by either MBNW protocol ($FRC_{CB}$ 3.17 (0.98) *versus* $FRC_{FB}$ 3.18 (0.94), p=0.88). $FRC_{CB}$ and $FRC_{FB}$ were strongly correlated (r=0.94, p<0.0001; figure 1a). There was no evidence of bias in the absolute (figure 1b) or percentage difference (supplementary figure S1).

| | Whole group | Repeatability subgroup |
|---|---|---|
| **TABLE 1** Participant characteristics, lung function and multiple-breath nitrogen washout (MBNW) parameters | | |
| **Females/males n** | 11/16 | 6/9 |
| **Age years median (range)** | 34 (19–65) | 30 (23–41) |
| **BMI kg·m$^{-2}$** | 24.6±3.4 | 25.1±4.2 |
| **Lung function** | | |
| FEV$_1$ % predicted[#] | 105±14 | 101±30 |
| FEV$_1$/FVC % | 83±6 | 84±5 |
| TLC$_{pleth}$ % predicted[¶] | 101±23 | 107±11 |
| FRC$_{pleth}$ % predicted[¶] | 97±27 | 104±20 |
| **MBNW parameters** | | |
| Controlled breathing | | |
| FRC$_{CB}$ L | 3.17±0.98 | 3.23±0.97 |
| LCI$_{CB}$ | 7.2±0.58 | 7.13±0.5 |
| $S_{cond_{CB}}$ L$^{-1}$ | 0.017±0.009 | 0.02±0.01 |
| $S_{acin_{CB}}$ L$^{-1}$ | 0.057±0.022 | 0.063±0.021 |
| $V_T$ mL | 1124±37 | 1127±37 |
| $V_T$/FRC$_{MBNW}$ | 0.40±0.15 | 0.39±0.14 |
| RR breaths·min$^{-1}$ | 12.21±1.36 | 11.9±1.53 |
| Free breathing | | |
| FRC$_{FB}$ L | 3.18±0.94 | 3.25±0.96 |
| LCI$_{FB}$ | 7.55±0.81 | 7.4±0.79 |
| $S_{cond_{FB}}$ L$^{-1}$ | 0.018±0.01 | 0.018±0.013 |
| $S_{acin_{FB}}$ L$^{-1}$ | 0.085±0.038 | 0.091±0.04 |
| $V_T$ mL | 880±325 | 912±304 |
| $V_T$/FRC$_{MBNW}$ | 0.29±0.11 | 0.3±0.13 |
| RR breaths·min$^{-1}$ | 13.78±3.43 | 13.18±3.46 |

Data are presented as mean±SD unless otherwise stated. For the repeatability subgroup, results are from the first visit. BMI: body mass index; FEV$_1$: forced expiratory volume in 1 s; FVC: forced vital capacity; TLC: total lung capacity; FRC: functional residual capacity; LCI: lung clearance index; $S_{cond}$: conductive zone ventilation heterogeneity; $S_{acin}$: acinar zone ventilation heterogeneity; $V_T$: mean tidal volume across measurements; RR: respiratory rate; pleth: body plethysmography; CB: controlled breathing protocol; FB: free breathing protocol. [#]: reference equations for predicted values from QUANJER et al. [13]; [¶]: reference equations for predicted values from QUANJER et al. [14].

There were no significant differences between mean FRC$_{CB}$ or FRC$_{FB}$ and mean FRC$_{pleth}$ (p=0.83 and p=0.86, respectively). Both FRC$_{CB}$ and FRC$_{FB}$ were strongly correlated with FRC$_{pleth}$ (r=0.84 and r=0.92, respectively, p<0.0001 for both; figure 2). Within-subject differences in FRC between the two MBNW protocols were not related to the individual's FRC$_{pleth}$ (p=0.55, supplementary figure S2-A). Similarly, within-subject differences in FRC between the protocols were not related to age, sex, height, mean $V_T$ or mean RR from free breathing, but there was a trend towards an effect of BMI (p=0.07, supplementary figure S2-B).

### LCI was comparable between protocols
Mean LCI was significantly lower with controlled breathing (LCI$_{CB}$ 7.2 (0.58) *versus* LCI$_{FB}$ 7.55 (0.81), p=0.0004); however, the mean difference was small. There was a strong correlation between the protocols (r=0.84, p<0.0001; figure 3a). There was evidence of proportional bias (between-protocol difference increased with LCI, p=0.004; figure 3b).

### $S_{cond}$ was poorly correlated between protocols
Overall, mean $S_{cond}$ measured by both protocols was not significantly different ($S_{cond_{CB}}$ 0.017 (0.009) *versus* $S_{cond_{FB}}$ 0.018 (0.01) L$^{-1}$, p=0.74). However, there was no significant correlation between the protocols (r=0.18, p=0.36, figure 4a). The Bland–Altman plot revealed large variance in between-protocol differences and possible (but non-significant) proportional bias (p=0.45, figure 4b).

### $S_{acin}$ showed significant between-protocol differences
Mean $S_{acin}$ was significantly lower with controlled breathing ($S_{acin_{CB}}$ 0.057 (0.021) L$^{-1}$ *versus* $S_{acin_{FB}}$ 0.085 (0.038) L$^{-1}$, p=0.0004). The correlation between the protocols was borderline significant (r=0.37, p=0.06; figure 5a).
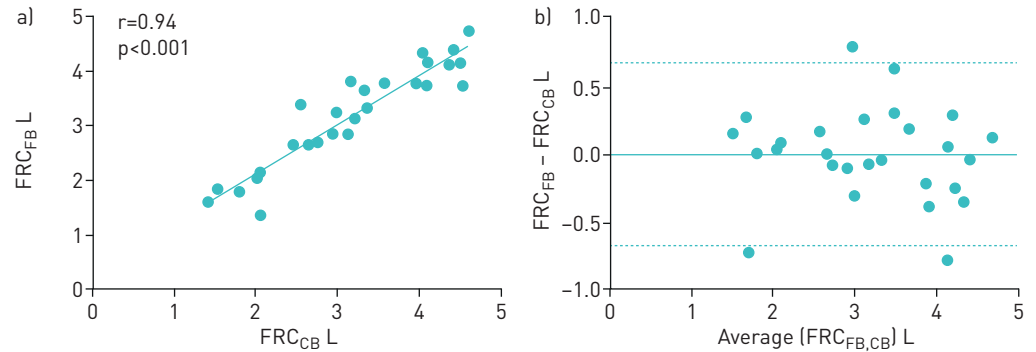
FIGURE 1 Functional residual capacity measured by controlled breathing ($FRC_{CB}$) and free breathing ($FRC_{FB}$) protocols. a) There was strong correlation between the protocols (r=0.94, p<0.0001). b) Bland–Altman plot showing good agreement between the protocols (mean difference (95% limits of agreement) 0.009 (−0.666, 0.686) L, p=0.88).

The Bland–Altman plot revealed significant proportional bias (between-protocol difference increased with $S_{acin}$, p=0.002; figure 5b). Linear regression showed that within-subject differences in $S_{acin}$ between the two protocols were related to the breathing pattern. Specifically, the greater the deviation in mean $V_T$ or mean RR between the protocols, the larger the discrepancy in $S_{acin}$ (p=0.004 and p=0.01, respectively; figure 5c and d), such that participants who breathed shallower or faster during the free breathing protocol had greater apparent $S_{acin_{FB}}$.

*Breathing protocol influenced between-session variability of MBNW indices*

Fifteen volunteers underwent repeat testing; the baseline (Visit 1) characteristics of this subgroup were similar to those of the group as a whole, except for a narrower age range (table 1). The mean (SD) time between sessions was 5.9 (3.3) weeks. Within-session and between-session variability measures for both protocols are presented in table 2. There were no differences seen in within-session CoV between protocols in FRC (p=0.677) or LCI (p=0.157). In terms of between-session variability, the free breathing protocol showed relatively greater variability in LCI, $S_{cond}$ and $S_{acin}$, as indicated by numerically greater mean differences and wider 95% LOAs, with very poor between-session ICC seen in $S_{cond}$ and $S_{acin}$. The controlled breathing protocol showed wider 95% LOA for FRC. Bland–Altman plots did not suggest any proportional bias for either protocol (supplementary figures S3 and S4).

## Discussion
### Summary of results
In this study in healthy adults comparing two commonly used MBNW breathing protocols, we found that: 1) FRC was comparable between the two protocols, however indices of ventilation heterogeneity based on phase III slopes ($S_{III}$) were not – $S_{cond}$ was poorly correlated between the two protocols and $S_{acin}$ was systematically higher under the free breathing protocol, whereas LCI was only marginally higher; 2) there was greater between-protocol discrepancy in $S_{acin}$ in subjects whose free-breathing pattern deviated from



FIGURE 2 Functional residual capacity measured by controlled breathing ($FRC_{CB}$) and free breathing ($FRC_{FB}$) protocols *versus* the gold-standard body plethysmography ($FRC_{pleth}$). There was good correlation between FRC measured by both protocols and $FRC_{pleth}$ (r=0.84 and r=0.92, respectively, p<0.0001 for both). MBNW: multiple breath nitrogen washout.
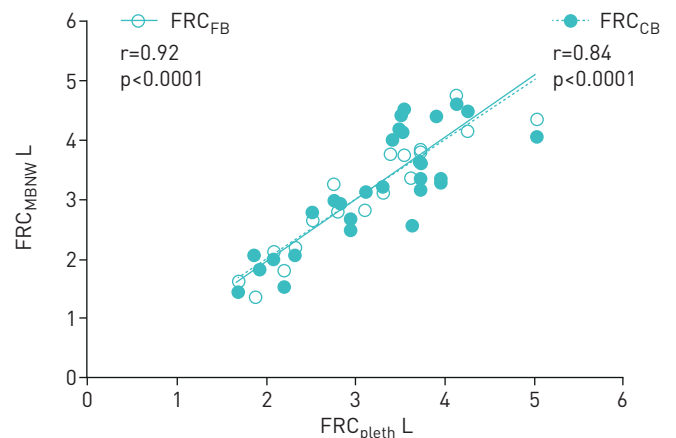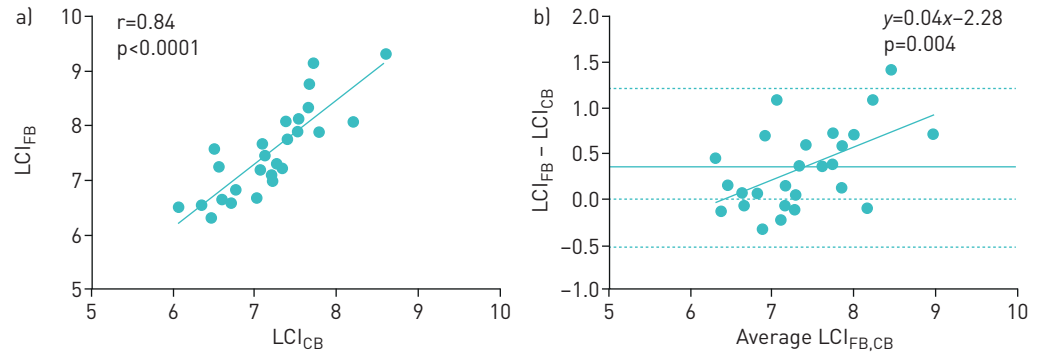
FIGURE 3 Lung clearance index measured by controlled breathing ($LCI_{CB}$) and free breathing ($LCI_{FB}$) protocols. a) There was strong correlation between the protocols (r=0.84, p<0.0001). b) Bland–Altman plot showing that free breathing produced a higher LCI compared to controlled breathing (mean difference (95% limits of agreement) 0.35 (−0.53, 1.23), p=0.0004)). There was also significant proportional bias confirmed by linear regression (p=0.004).

that of the volume- and frequency-controlled protocol; and 3) $S_{cond}$ and $S_{acin}$ exhibited poorer between-session repeatability under the free breathing protocol.

### Lack of discrepancy in FRC

YAMMINE et al. [12] demonstrated in children that $S_{cond}$ and LCI obtained from a controlled breathing protocol were higher than when obtained by free breathing and that the discrepancy was related to a decrease in FRC during controlled breathing. Thus, the authors argued that a controlled breathing protocol may overestimate heterogeneity by inadvertently causing individuals to change the EELV at which they would otherwise normally breathe. In children, a 1 L $V_T$ represents a significant proportion of total lung capacity; the high $V_T$ may increase ventilation heterogeneity due to the recruitment of normally non-ventilated lung units [19], or due to closure of lung units in the dependent regions if the child expires to below their normal EELV [20], or some combination of both.

Our current findings in healthy adults do not support the argument that FRC is altered with controlled breathing, as we showed that FRC estimated by the two protocols was in fact comparable, with no systematic bias. Furthermore, both were comparable to the "gold standard", i.e. $FRC_{pleth}$. A number of factors may explain the differences between our findings and those of YAMMINE et al. For example, the effect of variation in FRC may be relatively mild in the adults in our study compared to that in children, and thus insufficient to cause significant differences in heterogeneity. We observed that the variability in between-protocol differences (assessed by the LOA) in FRC were comparable to the between-session variabilities of either protocol, though interestingly between-session LOA was numerically higher under the controlled breathing protocol. Alternatively, our standard procedure for the controlled breathing protocol includes an initial period of unconstrained breathing to allow the individual's EELV to stabilise before
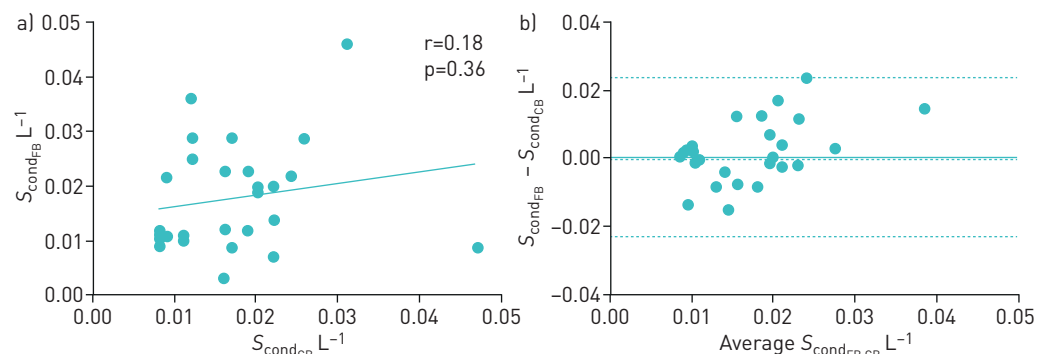


FIGURE 4 Ventilation heterogeneity in conducting airways measured by controlled breathing ($S_{cond_{CB}}$) and free breathing ($S_{cond_{FB}}$) protocols. a) There was no significant correlation between the protocols (r=0.18, p=0.36). b) Bland–Altman plot showing high between-protocol variability (mean difference (95% limits of agreement) 0.0008 (−0.02, 0.02) L$^{-1}$, p=0.74)). There appeared to be proportional bias on visual inspection, but linear regression was not statistically significant (p=0.45).
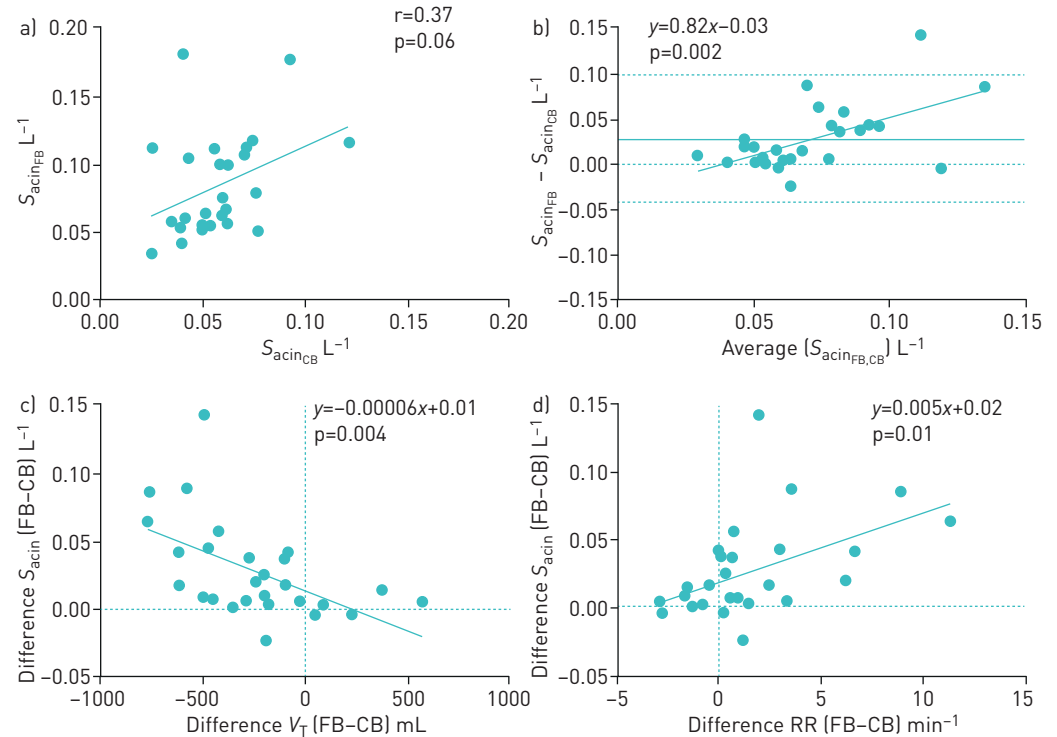
FIGURE 5 Ventilation heterogeneity in distal/intra-acinar airways measured by controlled breathing ($S_{acin_{CB}}$) and free breathing ($S_{acin_{FB}}$) protocols. a) There was relatively poor correlation between the protocols (r=0.37, p=0.06). b) Bland–Altman plot showing that free breathing produced higher $S_{acin}$ compared to controlled breathing (mean difference (95% limits of agreement) 0.03 (−0.04, 0.10) $L^{-1}$, p<0.0001)). There was also significant proportional bias confirmed by linear regression (p=0.002). The between-protocol difference in $S_{acin}$ ($S_{acin_{FB}}$–$S_{acin_{CB}}$) was predicted by the between-protocol differences in c) tidal volume ($V_{T_{FB}}$–$V_{T_{CB}}$, regression p=0.004) and d) respiratory rate ($RR_{FB}$–$RR_{CB}$, regression p=0.01). One participant was excluded from $S_{acin}$ analyses due to negative value in one trial.

commencing the washout. This is in line with current ATS/ERS recommendations [3] and may have ensured FRC differences were kept minimal between protocols.

### Discrepancy in ventilation heterogeneity indices

While LCI – an index of global heterogeneity that is independent of $S_{III}$ – was comparable between protocols (with a small bias), we found that $S_{III}$-dependent ventilation heterogeneity partitioned into proximal/conducting ($S_{cond}$) and distal/acinar ($S_{acin}$) zones was not. $S_{cond}$ was poorly correlated between protocols, perhaps driven by a large between-session variability. $S_{acin}$ was significantly different between protocols; moreover, $S_{acin}$ obtained by free breathing was higher than that obtained by controlled breathing in individuals with relatively small $V_T$ or high RR. Between-protocol differences in $S_{acin}$ appear to be independent of age, height or lung size, *i.e.* a change in breathing pattern was the dominant driver of the discrepancy.

There are two possible explanations for this. First, the effect could be purely methodological: when breaths are too "shallow", estimation of $S_{III}$ is rendered invalid as the $N_2$ expirogram either never reaches a plateau or attains a plateau that is too short for reliable curve fitting, which would subsequently affect the derivation of $S_{acin}$ and $S_{cond}$. Anecdotal evidence suggests this occurs frequently. Methods to more reliably estimate $S_{III}$ exist [21] but are still ultimately dependent on the presence of distinct phase II and III regions in the expirogram and an adequate portion of phase III. More useful may be methods of quantifying ventilation heterogeneity that do not rely on $S_{III}$ estimation [22, 23] – these may help not just in free breathing but also in pathologies where $S_{III}$ is often difficult to define. The feasibility and clinical relevance of these methods should be a direction for future investigation.

Alternatively, the effect may be physiological: when patients breathe shallower and faster, the fast-emptying lung compartments will increase their $N_2$ washout while the slow-emptying lung compartments will be unable to empty completely, thus increasing apparent heterogeneity estimated by $S_{acin}$ and $S_{cond}$. Indeed, the foundational studies on ventilation distribution clearly demonstrated an effect of breathing pattern

TABLE 2 Within- and between-session variability for the controlled and free breathing protocols

| | Within-session CoV | Between-session difference | 95% LOA | Between-session ICC |
|---|---|---|---|---|
| **Controlled breathing** | | | | |
| $FRC_{CB}$ L | 3.3±2.9% | −0.03±0.52 | −1.04–0.98 | 0.931 |
| $LCI_{CB}$ | 2.5±2.4% | 0.06±0.47 | −0.86–0.98 | 0.812 |
| $S_{cond_{CB}}$ $L^{-1}$ | | −0.001±0.011 | −0.017–0.015 | 0.836 |
| $S_{acin_{CB}}$ $L^{-1}$ | | −0.003±0.018 | −0.037–0.031 | 0.835 |
| **Free breathing** | | | | |
| $FRC_{FB}$ L | 3.6±2.3% | −0.05±0.29 | −0.62–0.52 | 0.980 |
| $LCI_{FB}$ | 3.2±1.5% | 0.15±0.53 | −0.89–1.19 | 0.850 |
| $S_{cond_{FB}}$ $L^{-1}$ | | 0.003±0.016 | −0.027–0.033 | 0.158 |
| $S_{acin_{FB}}$ $L^{-1}$ | | 0.004±0.052 | −0.098–0.106 | 0.334 |

Data are presented as mean±SD unless otherwise stated. Mean differences are Visit 2 minus Visit 1. CoV: coefficient of variation; 95% LOA: 95% limits of agreement; ICC: intra-class correlation coefficient; FRC: functional residual capacity; LCI: lung clearance index; $S_{cond}$: conducting airways ventilation heterogeneity; $S_{acin}$: distal/intra-acinar airways ventilation heterogeneity.

(including $V_T$) on MBNW indices [4]. Recently, RATJEN and colleagues [24] investigated the effects of altered $V_T$ on MBNW indices and found that shallow breathing (compared to "unrestricted" breathing) significantly increased LCI. The authors speculated that this was due to the effects of increased dead space to $V_T$ ratio on the determinants of LCI (both CEV and FRC). Notably, these investigators did not examine the impact on $S_{cond}$ or $S_{acin}$.

We found that, on average, $V_T$ was shallower, RR was faster and CEV higher in the free breathing protocol in our study participants. The change in CEV would be expected, since smaller breaths may reduce alveolar gas mixing efficiency and therefore increase the volume (and time, though this was not significant) required to reach end-of-test criteria. This likely drove the changes in LCI we observed – indeed, differences in LCI were directly proportional to differences in CEV (supplementary figure S5). However, it should be noted that the differences seen in LCI were very small relative to its magnitude and unlikely to be clinically significant. Furthermore, we found a weak relationship between differences in $S_{cond}$ and differences in CEV; this was not seen with $S_{acin}$ (supplementary figure S6). However, when plotted against differences in CEV/FRC (*i.e.* LCI, which corrects CEV for lung volume), the relationship with $S_{acin}$ became significant. This supports the interpretation that the differences seen in $S_{acin}$ were driven by changes in ventilation heterogeneity, rather than alterations to lung volume.

*Variability of MBNW indices*

We also report between-session variability over 2–10 weeks for both protocols, using the same commercially available device. This allowed us to compare between-protocol differences in MBNW indices against their short-term variability. Within-session CoV of FRC and LCI was similar to published values for the free breathing protocol [25], using the same device. However the between-session coefficient of repeatability (equal to 1.96×SD of the mean differences) of LCI, $S_{cond}$ and $S_{acin}$ was greater in our study, which may reflect the different time between sessions (weeks/months *versus* days). There is a paucity of repeatability data in health, and previous reports may not be generalisable due to persistent between-device differences [17, 18, 26]. Nevertheless, the excellent between-session repeatability for FRC and LCI [25, 27–29] but poor repeatability in $S_{acin}$ and $S_{cond}$ [25, 28] has been noted with the free breathing protocol by other investigators, regardless of device, tracer gas or disease. The higher between-session repeatability values seen in $S_{cond}$ and $S_{acin}$ with the controlled breathing protocol are consistent with other studies [30, 31].

It is also worth noting that FRC and LCI are derived from successive cumulation of volume at each breath, *i.e.* they are integrated measures, which are more robust to noise. On the other hand, $S_{cond}$ and $S_{acin}$ are estimated from slopes, *i.e.* differentiated measures, which are inherently susceptible to noise. Furthermore, unlike FRC and LCI, the values of $S_{cond}$ and $S_{acin}$ are very close to zero, which may explain the very low between-session ICC values observed. These may be fundamental reasons for the greater variability seen in $S_{cond}$ and $S_{acin}$ but not FRC and LCI, and this variability becomes even more pronounced when variability in tidal breathing is introduced. This higher variability may also drive the differences seen between the two protocols. We also note the potential for these differences to be further exaggerated in disease – this and

the high between-session variability observed may limit the clinical utility of $S_{III}$ indices derived from free breathing. These speculations need to be confirmed in further investigations.

### Significance

Both MBNW breathing protocols are designed to measure ventilation heterogeneity during resting tidal breathing. The controlled breathing protocol was originally devised based on modelling studies in which convection-dependent and diffusion-convection interaction-dependent mechanisms of airflow in the lung (the basis of $S_{cond}$ and $S_{acin}$, respectively) were first described [32]. The controlled breathing protocol standardises the $V_T$ at which ventilation distribution is assessed, and also ensures that expiration occurs well into the alveolar plateau so that a reliable SIII estimation can be made. This standardisation potentially minimises variability between tests. However, this "artificial" pattern of breathing may distribute the ventilation differently in lungs of different sizes, *i.e.* in shorter *versus* taller people. These differences may then affect the measurements of MBNW indices such that they no longer reflect true resting or "natural" ventilation distribution. In contrast, under a free breathing protocol, breathing occurs at (or closer to) natural resting breathing pattern; thus, the MBNW indices obtained arguably better reflect the individual's native ventilation distribution during normal tidal breathing.

Our results lend weight to the idea that controlled breathing imposes a condition that is different to the individual's natural resting breathing pattern, and that MBNW indices derived from this protocol may not necessarily reflect their "native" ventilation heterogeneity. Specifically, breathing at 1 L tended to improve heterogeneity measured by $S_{acin}$ in individuals whose $V_T$ during free breathing was <1 L, or whose RR during free breathing was faster than the prescribed 8–12 breaths per minute. So, which is the "better" test? We suggest that it is whichever provides the highest sensitivity in detecting disease or assessing treatment. The "better" test may even be disease-specific and warrants further investigation.

Since these differences are apparent despite the use of $V_T$-corrected versions of $S_{cond}$ and $S_{acin}$, our study suggests that the linear $V_T$ correction currently employed for the free breathing protocol may not be adequate. Again, methods which assess ventilation heterogeneity without the need for estimating $S_{III}$, ideally in a manner that is independent of $V_T$ [22], may be useful here.

### Limitations

Our study has several limitations to be taken into consideration. First, a number of the volunteers could be considered "trained" in lung function testing. However, it is unlikely that prior experience would influence the largely effort-independent MBNW test. Second, 4 out of 27 participants required coaching before or after free breathing testing due to insufficient expired volume for $S_{III}$ estimation. Thus, there is a chance our results may actually underestimate the true discrepancy between the two protocols, particularly in shallow breathers [24] or in lung disease where increased dead space further complicates $S_{III}$ estimation [33]. Third, our results only apply to a healthy population tested on a single MBNW device. There are likely to be differences between equipment [17, 18] and between populations (*e.g.* children, older adults, lung disease) that limit the generalisability of our results.

### Conclusion

We have demonstrated that while controlled and free breathing protocols are equivalent in terms of FRC and LCI, phase III-derived MBNW indices of ventilation heterogeneity are lower (*i.e.* better) when measured by the controlled breathing method. This effect seems to be driven largely by deviations in breathing pattern between the two protocols, particularly in patients who either breath faster or shallower than the criteria imposed by the controlled breathing protocol. Future work needs to be carried out in diseased patients to see if these observations hold true. Our study sheds light on potential physiological mechanisms behind these differences, and the overall interpretation of ventilation heterogeneity measured by MBNW. In better characterising the difference between the two protocols, these findings also help facilitate the ongoing efforts to standardise MBNW as an emerging clinical test.

## References

1   King GG. Cutting edge technologies in respiratory research: lung function testing. *Respirology* 2011; 16: 883–890.

2   Zimmermann SC, Tonga KO, Thamrin C. Dismantling airway disease with the use of new pulmonary function indices. *Eur Respir Rev* 2019; 28: 180122.

3   Robinson PD, Latzin P, Verbanck S, *et al.* Consensus statement for inert gas washout measurement using multiple- and single-breath tests. *Eur Respir J* 2013; 41: 507–522.

4   Crawford AB, Makowska M, Engel LA. Effect of tidal volume on ventilation maldistribution. *Respir Physiol* 1986; 66: 11–25.

5   Jones JG. The effect of preinspiratory lung volume on the result of the single breath O2 test. *Respir Physiol* 1967; 2: 375–385.

6   Gronkvist M, Bergsten E, Gustafsson PM. Effects of body posture and tidal volume on inter- and intraregional ventilation distribution in healthy men. *J Appl Physiol (1985)* 2002; 92: 634–642.

7   Aurora P, Kozlowska W, Stocks J. Gas mixing efficiency from birth to adulthood measured by multiple-breath washout. *Respir Physiol Neurobiol* 2005; 148: 125–139.

8   Anagnostopoulou P, Latzin P, Jensen R, *et al.* Normative data for multiple breath washout outcomes in school-aged Caucasian children. *Eur Respir J* 2020; 55: 1901302.

9   Kjellberg S, Houltz BK, Zetterstrom O, *et al.* Clinical characteristics of adult asthma associated with small airway dysfunction. *Respir Med* 2016; 117: 92–102.

10  Verbanck S, Thompson BR, Schuermans D, *et al.* Ventilation heterogeneity in the acinar and conductive zones of the normal ageing lung. *Thorax* 2012; 67: 789–795.

11  Verbanck S, Van Muylem A, Schuermans D, *et al.* Transfer factor, lung volumes, resistance and ventilation distribution in healthy adults. *Eur Respir J* 2016; 47: 166–176.

12  Yammine S, Singer F, Gustafsson P, *et al.* Impact of different breathing protocols on multiple-breath washout outcomes in children. *J Cyst Fibros* 2014; 13: 190–197.

13  Quanjer PH, Stanojevic S, Cole TJ, *et al.* Multi-ethnic reference values for spirometry for the 3–95-yr age range: the global lung function 2012 equations. *Eur Respir J* 2012; 40: 1324–1343.

14  Quanjer PH, Tammeling GJ, Cotes JE, *et al.* Lung volumes and forced ventilatory flows. *Eur Respir J* 1993; 6: Suppl. 16, 5–40.

15  Verbanck S, Schuermans D, Van Muylem A, *et al.* Ventilation distribution during histamine provocation. *J Appl Physiol (1985)* 1997; 83: 1907–1916.

16  Koo TK, Li MY. A guideline of selecting and reporting intraclass correlation coefficients for reliability research. *J Chiropr Med* 2016; 15: 155–163.

17  Tonga KO, Robinson PD, Farah CS, *et al.* In vitro and in vivo functional residual capacity comparisons between multiple-breath nitrogen washout devices. *ERJ Open Res* 2017; 3: 00011-2017.

18  Poncin W, Singer F, Aubriot A-S, *et al.* Agreement between multiple-breath nitrogen washout systems in children and adults. *J Cyst Fibros* 2017; 16: 258–266.

19  Christensson P, Arborelius M, Jr., Kautto R. Volume of trapped gas in lungs of healthy humans. *J Appl Physiol Respir Environ Exerc Physiol* 1981; 51: 172–175.

20  Edward JBJ, Macklem P. Airway closure: demonstration by breathing 100% $O_2$ at low lung volumes and by $N_2$ washout. *J Appl Physiol* 1968; 25: 139–148.

21  Stuart-Andrews CR, Kelly VJ, Sands SA, *et al.* Automated detection of the phase III slope during inert gas washout testing. *J Appl Physiol (1985)* 2012; 112: 1073–1081.

22  Bates JHT, Peters U. A model-based approach to interpreting multibreath nitrogen washout data. *J Appl Physiol (1985)* 2018; 124: 1155–1163.

23  Lewis SM, Evans JW, Jalowayski AA. Continuous distributions of specific ventilation recovered from inert gas washout. *J Appl Physiol Respir Environ Exerc Physiol* 1978; 44: 416–423.

24  Ratjen F, Jensen R, Klingel M, *et al.* Effect of changes in tidal volume on multiple breath washout outcomes. *PLoS One* 2019; 14: e0219309.

25  Singer F, Houltz B, Latzin P, *et al.* A realistic validation study of a new nitrogen multiple-breath washout system. *PLoS One* 2012; 7: e36083.

26  Zwitserloot AM, van den Born EJ, Raaijmakers LHA, *et al.* Differences in lung clearance index and functional residual capacity between two commercial multiple-breath nitrogen washout devices in healthy children and adults. *ERJ Open Res* 2020; 6: 00247-02019.

27  Bell AS, Lawrence PJ, Singh D, *et al.* Feasibility and challenges of using multiple breath washout in COPD. *Int J Chron Obstruct Pulmon Dis* 2018; 13: 2113–2119.

28  Horsley AR, Macleod KA, Robson AG, *et al.* Effects of cystic fibrosis lung disease on gas mixing indices derived from alveolar slope analysis. *Respir Physiol Neurobiol* 2008; 162: 197–203.

29  Isaac SM, Jensen R, Anagnostopoulou P, *et al.* Evaluation of a multiple breath nitrogen washout system in children. *Pediatr Pulmonol* 2020; 55: 2108–2114.

30  Gonem S, Corkill S, Singapuri A, *et al.* Between-visit variability of small airway obstruction markers in patients with asthma. *Eur Respir J* 2014; 44: 242–244.

31  Jetmalani K, Chapman DG, Thamrin C, *et al.* Bronchodilator responsiveness of peripheral airways in smokers with normal spirometry. *Respirology* 2016; 21: 1270–1276.

32  Verbanck S, Paiva M. Model simulations of gas mixing and ventilation distribution in the human lung. *J Appl Physiol (1985)* 1990; 69: 2269–2279.

33  Fowler WS. Lung function studies. III. Uneven pulmonary ventilation in normal subjects and in patients with pulmonary disease. *J Appl Physiol* 1949; 2: 283–299.