

Reply to Migration is not the perfect answer: Optimised methodology to assess LCI agreement between corrected legacy multiple breath nitrogen washout data and that directly collected on updated software.

Paul Robinson¹, Ediane De Queiroz Andrade¹, Brooke Bailey¹, Jane Davies², Renee Jensen³, Felix Ratjen³, Clare Saunders², and Christopher Short²

¹The Children's Hospital at Westmead

²Royal Brompton and Harefield Hospitals

³The Hospital for Sick Children

March 30, 2023

Title: Reply to Migration is not the perfect answer: Optimised methodology to assess LCI agreement between corrected legacy multiple breath nitrogen washout data and that directly collected on updated software.

Authors:

Ediane De Queiroz Andrade^{1,2}

Brooke Bailey¹

Jane C. Davies^{3,4,5}

Renee Jensen^{6,7,8}

Felix Ratjen^{6,7,8,9}

Clare J. Saunders^{3,4,5}

Christopher Short^{3,4,5}

Paul D Robinson^{1,2,10,11}

Affiliations:

¹Department of Respiratory Medicine, The Children's Hospital at Westmead, Sydney, New South Wales, Australia;

²Australian Multiple Breath Washout Central Over-Reading Centre, Sydney, Australia

³Royal Brompton and Harefield Hospitals, Guys and St Thomas' Trust, London, UK;

⁴National Heart and Lung Institute, Imperial College London, London, UK;

⁵European Cystic Fibrosis Society Lung Clearance Index Core Facility, London, United Kingdom

⁶Translational Medicine, Hospital for Sick Children, Toronto, Ontario, Canada;

⁷Division of Respiratory Medicine, Hospital for Sick Children, Toronto, Ontario, Canada;

⁸Cystic Fibrosis Foundation Multiple Breath Washout National Resource Centre, Toronto, Canada

⁹University of Toronto, Toronto, Canada;

¹⁰Discipline of Paediatrics and Child Health, University of Sydney, Sydney, New South Wales, Australia;

¹¹Airway Physiology and Imaging group, Woolcock Medical Research Institute, Sydney, New South Wales, Australia;

Word count: 987 (up to 1000 words permitted)

CORRESPONDING AUTHOR

Professor Paul D Robinson

The Children’s Hospital at Westmead The Sydney Children’s Hospital Network Respiratory Medicine Department Cnr Hawkesbury Road and Hainsworth Street Locked Bag 4001, Westmead 2145, NSW Australia
t: (02) 9845 3395/97 | f: (02) 9845 3396 | e: dr.pdrobinson@gmail.com

To the Editor,

The recent identification of a sensor crosstalk (XT) error in the Exhalyser D MBW device¹ led to the release of new software (3.3.1). Understanding impacts of transition between MBWN₂ software updates, and how to best handle historical data, is critical². Should users simply migrate legacy data, by uploading into the new software version to enable XT correction (XTC), or perform a more time-consuming reload of raw data A-files to apply *all* advances between that and historical software versions (3.1.6): XTC, O₂ drift correction and novel gas signal flow synchronisation methodology (dynamic delay correction, DDC)? What are the differences and which approach most accurately reflects MBWN₂ Lung Clearance Index (LCI) values collected within 3.3.1?

This issue has been investigated to varying degrees across three studies published in 2022. Two studies have drawn differing conclusions comparing agreement between migration and reloading. Oestreich et al³ (n=44, healthy and Cystic Fibrosis, CF) reported a mean (95% CI) LCI difference of -0.16 (-0.27 to -0.04) turnovers (TO) or -1.3%, with an observed magnitude-dependent bias, which they felt supported a recommendation to reload data (and not migrate). Jensen et al.⁴, in a reply to that letter, “respectfully disagreed”. In their independent analysis of a separate, similarly sized cohort, in which they attempted to isolate the impact of XTC on 3.1.6 data, mean (95% CI) difference between migration and reload was -0.10 (-0.26, 0.06), or -1%, was not statistically significant, without a magnitude-dependent bias when correctly expressed as relative difference.

In the next study to be published, by Short et al, this discussion was extended by introducing *direct* comparison to 3.3.1 collected data. Subjects (n=19, healthy and CF) performed [?]2 technically acceptable trials on both 3.1.6 and 3.3.1 in a fixed order (mandated by study protocol). They confirmed the non-significant difference between migrated and reloaded data reported by Jensen, but also showed a large difference between collected 3.3.1 data and 3.1.6-collected data ‘corrected’ via either migration or reloading into 3.3.1: mean absolute difference 0.5-0.8, relative difference 7.2-8.5%. The authors felt these differences raised concerns about ‘corrected’ 3.1.6 data being used by the current GLI normative values project to derive reference equations applicable for 3.3.1, and proposed that, ideally, only data collected in 3.3.1 should be included. This conclusion would have significant implications on the available data to derive these equations.

Here we present results from an independent Australian dataset that the 3 CORCs have scrutinised and agree offers methodological improvements to help maximise the utility of legacy data. In this study, performed under local ethics approval (2021/ETH11854), recruited participants performed [?]2 technically acceptable trials in both 3.1.6 and 3.3.1, in a randomised order, on a single Exhalyser D device at a single visit. 3.1.6 data were both migrated and reloaded in 3.3.1 to generate ‘corrected’ LCI values to compare to data collected directly on version 3.3.1. Aspects of processing within 3.3.1 were identical for reloaded and 3.3.1 collected data, as done by Short et al, but with the addition of consistent O₂ and CO₂ sensor offset values. The latter

was not performed in the study by Short et al where instead, software and equipment specific offsets were used to produce effective synchronisation (assessed visually) during the washout phase.

Summary of signal processes applied:

	Migrated	Reloaded	Collected
Fixed signal offset		x	x
DDC	x	*	*
O ₂ drift correction	x		
XTC			

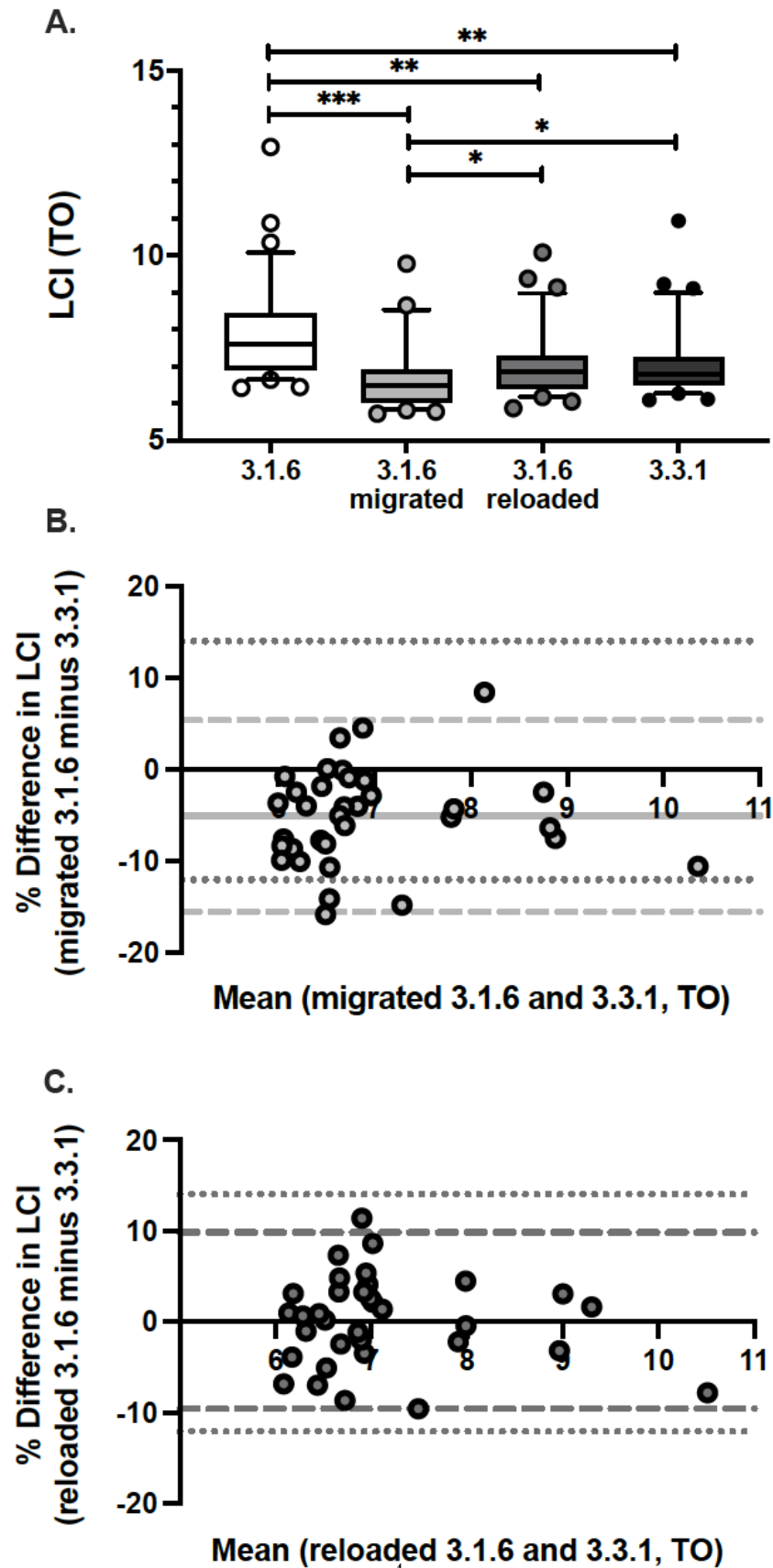
*Identical offset values for signal alignment applied to reloaded and collected data

Kruskal–Wallis nonparametric tests compared LCI values. As differences between the three methods were normally distributed, Bland Altman analysis assessed agreement, and as an additional comparison 95% limits of agreement (LA) were compared to within-session repeatability between the two 3.3.1 collected trials.

All 34 participants contributed complete sets of technically acceptable data: mean (SD; range) age 15 (9;5-47) years, 28 with CF. LCI agreement is summarised in Figure 1. In agreement with Short et al, there was a difference between migrated values vs 3.3.1 collected LCI, albeit smaller in magnitude: 6.50(5.72-9.78) vs 6.79(6.09-10.94), $p=0.049$; mean (95% LA) difference -0.37 (-1.17-0.66). But unlike in that study, the additional optimisation of methodology resulted in reloaded values which were not significantly different from those directly collected in 3.3.1: 6.86(5.87-10.09), $p=0.819$; actual difference 0.01(-0.85,0.75) and 0.13(-9.57,9.83)% change. Differences to both migrated and reloaded 3.1.6 data were almost entirely within this study’s within-session repeatability of 3.3.1 collected data [mean 95% LA difference 0.11(-0.83,1.05) and -0.98(-12.0,14.0%), respectively]. We did find evidence of a within-subject software version sequence effect on LCI difference within this dataset, in the same direction of the difference described by Short et al, supporting the randomisation approach taken in this study’s methodology (data not shown).

Collectively the CORCs believe these data together with the three previous datasets³⁻⁵ are now sufficient to make some recommendations to the community. Firstly, that reloaded legacy 3.1.6 data agree strongly with contemporaneously collected 3.3.1 data when study design carefully ensures alignment of all the signal processing factors previously discussed, including ensuring consistent offset values across comparison groups. Secondly, all these studies have demonstrated that, whilst data migration does generate LCI values that differ from prospectively collected 3.3.1 data (due to the lack of O₂ drift correction and DDC application), magnitude of difference is relatively small (<0.4 TO, or 5%) and usually within the within-test repeatability. Depending on the reasons for researchers wishing to use longitudinal, legacy data, these differences may be deemed acceptable. Thirdly, variation in findings across these studies highlights that the methodology used to update legacy data can impact results due to varying application of signal processing factors, and care is needed when updating legacy data (whether by migration or reloading), to ensure the method used is fit for purpose, and the process used must be fully transparent and described. The strong agreement with optimised, reloaded data within this study is reassuring for the current GLI normative data project which now specifies that legacy data is reloaded by contributing sites using a detailed protocol to minimise between site differences in approach. As the amount of 3.3.1 collected data contributed to the GLI database increases, the ability to evaluate agreement with legacy reloaded 3.1.6 data will be created and should be evaluated.

Figure 1. Agreement between MBWN2 data: Collected and processed in 3.1.6, later migrated to 3.3.1, reloaded in 3.3.1 in comparison to data collected and processed in 3.3.1 (panel A). Bland Altman plots outlining the percent difference between data collected in 3.1.6 and migrated to (panel B) or reloaded in (panel C) 3.3.1, and data collected in 3.3.1, in comparison to within-test repeatability of 3.3.1 collected data.



Footnote: * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$ vs. 3.3.1 collected values. Band Altman plots show mean difference as solid horizontal lines, 95% limits of agreement (95% LA) as wide dashed horizontal lines, and comparison 95% LA values for within-session repeatability of 3.3.1 collected data as narrow dashed lines.

References

1. Wyler F, Oestreich MA, Frauchiger BS, Ramsey KA, Latzin P. Correction of sensor crosstalk error in Exhalyzer D multiple-breath washout device significantly impacts outcomes in children with cystic fibrosis. *Journal of applied physiology* 2021;131(3):1148-1156.
2. Robinson PD, Latzin P, Verbanck S, Hall GL, Horsley A, Gappa M, Thamrin C, Arets HG, Aurora P, Fuchs S, *et al* . ERS/ATS Consensus statement for inert gas washout measurement using multiple and single breath tests. *Eur Respir J* 2013;41(3):507-522.
3. Oestreich MA, Wyler F, Latzin P, Ramsey KA. Impact of Spiroware re-analysis method on multiple-breath washout outcomes in children with cystic fibrosis. *J Cyst Fibros* 2022;21(3):e208-e209.
4. Jensen R, Shaw M, Phanthanourak A, Ratjen F. Re: Impact of spiroware re-analysis method on multiple-breath washout outcomes in children with cystic fibrosis; M.A. Oestreich, F. Wyler, P. Latzin et al. *J Cyst Fibros* 2022;21(3):e210-e211.
5. Short C, Abkir M, Pinnell S, Proctor O, Saunders CJ, Davies JC. Migration is not the perfect answer: How the cross-talk error correction for multiple breath nitrogen washout (MBWN(2)) parameters differs on directly collected vs. legacy data. *Pediatr Pulmonol* 2023;58(1):328-331.