

Lung Clearance Index as a definitive surrogate endpoint for therapeutic interventions in cystic fibrosis

Katie J. Bayfield

Imperial College London, National Heart and Lung Institute

Thesis submitted for Doctor of Philosophy June 2017

Word count: 77,346 (Plus References 7010, Appendices 2178)

0.1 Declaration of originality

I declare that the work described in this thesis is my own original work, and that any work undertaken by others has been appropriately attributed and referenced in the text.

0.2 Declaration of copyright

The copyright of this thesis rests with the author and is made available under a Creative Commons Attribution Non-Commercial No Derivatives license. Researchers are free to copy, distribute or transmit the thesis on the condition that they attribute it, that they do not use it for commercial purposes and that they do not alter, transform or build upon it. For any reuse or redistribution, researchers must make clear to others the license terms of this work.

0.3 Abstract

Lung Clearance Index (LCI), a measure of ventilation inhomogeneity, reflects disease in distal airways predominantly affected in cystic fibrosis (CF). I hypothesised that improved understanding of LCI's clinimetric properties could support its progression to an accepted outcome measure. I aimed to define and explain differences between techniques and assess responsiveness to interventions.

LCI is derived from multibreath washout of inert gas, usually sulphur hexafluoride (SF₆) or Nitrogen (N₂). I tested 47 CF patients and 42 healthy controls on both the Exhalyzer D (N₂) and a modified Innocor device (SF₆) demonstrating that although both correlated with forced expiratory volume in the 1st second (FEV₁), LCI from N₂ was consistently higher i.e. more abnormal. I designed a novel simultaneous washout system, which indicated that tissue N₂ was likely a major contributor. I produced normal range data and confirmed the absence of diurnal effects.

I was the senior site physiologist for the UK CF Gene Therapy Consortium's phase 2b multi-dose trial. 116 CF patients completed 12 months of nebulised liposome-mediated gene therapy or placebo; the primary outcome, relative improvement in FEV₁, was met. Similar improvements were not detected by LCI. My detailed exploratory analysis failed to fully explain this although it seems likely related to failure of sufficient inhaled drug reaching the distal airways.

In parallel, I performed serial LCIs on CF patients prescribed treatment changes. Although FEV₁ improved LCI did not, even in subgroups based on disease severity/ stability and intervention. Sample sizes were small, and patient convenience precluded standardisation of time points; further prospective data are needed.

This thesis contributes to our understanding of differences between LCI techniques, emphasising the fact that results are not interchangeable. Unlike in early CF, when used in a population with wider disease severity, LCI does not appear to offer an advantage for monitoring over FEV₁.

0.4 Acknowledgements

Thank you to Professor Jane Davies, Professor Eric Alton and Clare Saunders for welcoming me into the clinical research field and inviting me to be part of a truly remarkable team. Together with you and the clinical team during the gene therapy multi-dose trial, I not only learnt about clinical trials and translational medicine but I also had a “work family”, many of which will remain close friends. With all of your help I grew as a researcher but also as a person and I will always value my time here.

My PhD and time at Imperial would not have been possible with Jane, thank you for seeing the potential in me, for putting up with my idiosyncrasies, laughing with me and having a shoulder to cry on. You are truly an inspiration and if I could have a thimble of your knowledge, poise, passion and aptitude for research, then I could also one day be an admirable role model.

Particular thanks should go to Samantha Irving for her teaching and encouragement. I am honoured to have received honest and valuable advice from Professor Andy Bush. I am very grateful for input and experience from Dr Alex Horsley. I would also like to thank expert researchers and colleagues for their advice and direction when speaking together at various conferences.

I am privileged to have worked with many people at the Royal Brompton Hospital, my Mentor, Simon Ward and the RBH lung function department as well as the adult CF and paediatric respiratory departments. Their expertise knows no bounds and I am pleased to have been welcomed into the multidisciplinary team. They also helped me to connect with the cystic fibrosis patients and their families, of which, this thesis would not have been possible. I will forever be indebted to those patients who dedicated their time and attention into helping me; your altruism and positivity is honourable.

Thank you to my friends at Imperial who have endured silly questions, kept me sane and offered light lunchtime relief.

My “Cheerleaders” outside of work have listened, offered words of wisdom, exercised with me and given comfort with their infallible friendship and ability to make me smile.

Finally thank you to my Mum and Dad, who have always encouraged and believed I could achieve anything I set my mind to and to my brother, with whom I have never been closer. I hope I have made you all proud.

0.5 Abbreviations

ABPA	Allergic Bronchopulmonary Aspergillosis
ANCOVA	Analysis of covariance
ANOVA	Analysis of variance
ASCII	American Standard Code for Information Interchange
ATPC	Automatic Total Pressure Correction
ATS	American Thoracic Society
AZLI	Astreonam
BTPS	Body temperature and pressure, saturated with water
cAMP	Cyclic adenosine monophosphate
CDI	Convection dependant inhomogeneity
Cet	End tidal gas concentration
CEV	Cumulative expiratory volume
CF	Cystic Fibrosis
CFTR	Cystic Fibrosis Transmembrane Regulator protein
CFQR QoL	Cystic Fibrosis Questionnaire Revised, Quality of Life questionnaire (respiratory domain)
CO ₂	Carbon Dioxide
CT	Computed Tomography
CTN	Clinical trials network
Curv	Curvilinearity
CV%	Coefficient of Variation in percentage
DCDI	Diffusion-convection dependant inhomogeneity
DO1	DO12 – Dose 1 – Dose 12
DNA	Deoxyribonucleic acid
DNase	Dornase alpha
DSR	Deadspace reducer
ECFS	European cystic fibrosis society
EME	Efficacy and Mechanism Evaluation
ENaC	Epithelial sodium channel
ERS	European Respiratory Society
ERV	Expiratory Reserve Volume
ExD	Exhalyzer D, Ecomedics AG, Switzerland,
ExD/N ₂	Exhalyzer D, Ecomedics AG, Switzerland, N ₂ as a tracer gas

Exp	Expiration
FDA	Food and Drug Administration
FEV ₁	Forced Expiratory Volumes in 1 second
FEV ₁ %	Forced Expiratory Volumes in 1 second, percentage
predicted	
FGD	Flow gas delay
FRC	Functional residual capacity
FU1 and FU2	Post-dosing follow up visits
FVC	Forced vital capacity
GLI	Global Lung Initiative
GL67A	Cationic lipid combination
g/L	Grams per litre
g/mol	Grams per mole
GTAC	Gene Therapy advisory committee
GTC	United Kingdom Cystic Fibrosis Gene Therapy Consortium
HC	Healthy control
He	Helium
Hz	Hertz
IC	Inspiratory capacity
ICTU	Imperial Clinical Trials Unit
IgE	Immunoglobulin
Inn/SF ₆	Innocor™ gas analyser (Innovision, Denmark), SF ₆ as a tracer
gas	
Ino	Innocor gas analyser (Innovision, Denmark),
Insp	Inspiration
INT	Introduction visit
ITT	Intention to treat
IV	Intravenous (antibiotics)
IVC	Inspiratory vital capacity
kg	Kilograms
K	Kelvin
LCI	Lung Clearance Index
LCI _{2.5}	Lung Clearance Index calculated at the 1/40 th end tidal gas
concentration	

LCI ₅	Lung Clearance Index calculated at the shorter 1/20 th end
tidal gas concentration	
Mac	Macrophage
MBW	Multiple Breath Washout
MCID	Minimal clinically important difference
MDT	United Kingdom Cystic Fibrosis Gene Therapy Multi-dose
trial	
MEF ₂₅ , MEF ₅₀ , MEF ₂₅₋₇₅ , MEF ₇₅	Mid Expiratory flow between different percentages (25, 50, 25-75 and 75%) of the breath
mg	Milligram
MHRA	Medicines and healthcare product regulatory agency
ml	Millilitres
mm	Millimetres
mmHG	Millimetres of Mercury
mmol/L	Millimoles per litre
MRI	Magnetic Resonance Imaging
mRNA	Messenger ribonucleic acid
ms	Milliseconds
NDD	New diagnostic design technology – EasyOne Pro device
NHS	National Health Service
NIHR	National Institute of Health Research
NPD	Nasal potential difference
NE	Neutrophil elastase
N ₂	Nitrogen
N ₂ O	Nitric oxide
ns	Non significant
O ₂	Oxygen
Pa	Pascal
pGM169	Plasmid DNA
PhD	Doctor of Philosophy
Pleth	Plethysmography
pmn	Neutrophil
PP	Per protocol
RBH	Royal Brompton Hospital

RMS	Respiratory Mass Spectrometry
RR	Respiratory Rate
RV	Residual volume
Sacin (*) <i>al.</i> new cut offs)	Phase III slope analysis in the acinar region (* = Verbanck <i>et</i>
Scond (*) Verbanck <i>et al.</i> new cut offs)	Phase III slope analysis in the conductive region (* =
SCR	Screening visit
SD	Standard deviation
SF ₆	Sulphur hexafluoride
SOP	Standard operating procedure
SW	Simple Washout software
TLC	Total lung capacity
UNK	Unknown
ULN	Upper limit of normal
VI	Ventilation Inhomogeneity
V _T	Tidal volume

0.6 Publications and Presentations arising

Journal articles

Saunders, C., **Bayfield, K.**, Irving, I., Short, C., Bush, A., Davies, J. *Developments in multibreath washout testing in children with cystic fibrosis*. Curr Med Res Opin, 2016: p. 1-18.

This review article was completed in 2016, I compiled the MBW background section and the technical aspects of MBW testing.

Bayfield, K.J., Saunders, C, Alton, E.W., Davies, J.C. *Lung clearance Index and the importance of standardisation within clinical trials*. Inspire, The journal of the Association for Respiratory Technology & Physiology, 2015. **16**(3): p. 32-36.

This article was completed for my physiological society (Association for Respiratory Technology & Physiology - ARTP) after receipt of an ERS travel grant award from the ARTP 2015.

Alton, E.W., Armstrong, D. K., Ashby, D., **Bayfield, K. J.**, Bilton, D., Bloomfield, E. V., Boyd, A. C., Brand, J., Buchan, R., Calcedo, R., Carvelli, P., Chan, M., Cheng, S. H., Collie, D. D., Cunningham, S., Davidson, H. E., Davies, G., Davies, J. C., Davies, L. A., Dewar, M. H., Doherty, A., Donovan, J., Dwyer, N. S., Elgmati, H. I., Featherstone, R. F., Gavino, J., Gea-Sorli, S., Geddes, D. M., Gibson, J. S., Gill, D. R., Greening, A. P., Griesenbach, U., Hansell, D. M., Harman, K., Higgins, T. E., Hodges, S. L., Hyde, S. C., Hyndman, L., Innes, J. A., Jacob, J., Jones, N., Keogh, B. F., Limberis, M. P., Lloyd-Evans, P., Maclean, A. W., Manvell, M. C., McCormick, D., McGovern, M., McLachlan, G., Meng, C., Montero, M. A., Milligan, H., Moyce, L. J., Murray, G. D., Nicholson, A. G., Osadolor, T., Parra-Leiton, J., Porteous, D. J., Pringle, I. A., Punch, E. K., Pytel, K. M., Quittner, A. L., Rivellini, G., Saunders, C. J., Scheule, R. K., Sheard, S., Simmonds, N. J., Smith, K., Smith, S. N., Soussi, N., Soussi, S., Spearing, E. J., Stevenson, B. J., Sumner-Jones, S. G., Turkila, M., Ureta, R. P., Waller, M. D., Wasowicz, M. Y., Wilson, J. M., Wolstenholme-Hogg, P. et al., *Repeated nebulisation of non-viral CFTR gene therapy in patients with cystic fibrosis: a randomised, double-blind, placebo-controlled, phase 2b trial*. Lancet Respir Med, 2015.

This is the main paper providing the results for the UK CF gene therapy consortium multi-dose trial. I was the lead physiologist on the trial from the Royal Brompton Hospital and Imperial College site. I performed tests at trial visits, gathered and compiled together the data for data analysis. Reviewed paper and provided confirmation of trial subjects, contributed to the explanation text in the supplement.

Conference abstracts and presentations

Bayfield, K., Alton, EFWF, Horsley, A, Davies, J.C., *Further comparison of lung clearance index (LCI) generated with SF6 and N2-based multiple breath washout (MBW) techniques: differences increase with disease severity and are predominantly related to the prolonged final stage of the N2 washout*. Journal of Cystic Fibrosis, 2016. **15**(Supplement): p. S24.

I was awarded the European CF Society Young Investigator Award for this work, I presented at the ECFS conference (2016, Basel) in a session entitled "New lung function methods to monitor disease and treatment".

Bayfield, K., Alton, E., Irving, S., Bush, A., Horsley, A., Davies, J., *Simultaneous SF6 and N2 gas multiple breath washout (MBW); understanding the difference between test gases.* European Respiratory Journal, 2016. **48** (Supplement 60): p. 285.

I had an oral presentation of this work at the 2016 ERS in London, in a session entitled "The best is yet to come in lung function".

Bayfield, K, J., Alton EFWF., Irving, S., Bush, A., Horsley, A,R., Davies, J,C. *Simultaneous SF6 and N2 gas Multiple Breath Washout (MBW); understanding the difference between test gases- Abstract Reviews.* European Medical Journal, 2016, Respiratory, p. 67-68.

This was a review of the 2016 ERS session and presentation above.

Bayfield, K., Alton, E., Irving, S., Davies, J., *Comparison of Functional Residual Capacity (FRC) from two multiple breath washout (MBW) systems and Body Plethysmography.* European Respiratory Journal, 2016. **48**(Supplement 60).

This was presented in an spoken e poster session at the ERS 2016 London.

Saunders, C., **Bayfield, K.,** et al., *134 Training and qualifying international teams in standardised procedures: steps on the learning curve from the CTN LCI Core Facility.* Journal of Cystic Fibrosis, 2016. **15**: p. S85.

This work was presented in poster format at the ECFS 2016 (London). I was a physiologist assisting with training as part of the LCI core facility team.

Bayfield, K., et al., *P79 Comparison of CF and Non CF LCI results using the Exhalyzer D and Innocor™ devices.* Thorax, 2015. **70**(Suppl 3): p. A115.

This work was presented in an e poster session at the 2015 British Thoracic Society meeting.

Bayfield, K.J., et al., *ePS04.3 Comparison of CF and non-CF LCI results using the Exhalyzer D and Innocor™ devices.* Journal of Cystic Fibrosis, 2015. **14, Supplement 1**: p. S48.

This work was presented in an e poster session at the 2015 ECFS conference (Brussels).

Saunders, C., **Bayfield, K.** et al., *WS21.9 Lung clearance index as a clinical trial outcome measure: Establishing the CTN LCI core facility.* Journal of Cystic Fibrosis, 2015. **14, Supplement 1**: p. S41.

This was presented at the ECFS 2015 conference (Brussels). I helped to set up and establish the LCI core facility at the London site.

Bayfield, K.J., et al. *Lung function outcome measures in a multidose trial of non-viral gene therapy in patients with CF.* Poster Session Abstracts, 410. Pediatric Pulmonology, 2014. **49**(S38): p. S216-S456.

This work was presented in poster format at the North American CF Conference in 2014 (Atlanta).

Elgmati, H. I., Alton, E., Armstrong, D., **Bayfield, K.** et al. *Cardiopulmonary Exercise testing (CPET) as an outcome measure in cystic fibrosis clinical trial*. Poster Session Abstracts, 413. Pediatric Pulmonology, 2014. **49**(S38): p. S216-S456.

This work was presented in poster format at the North American CF Conference in 2014 (Atlanta). I was the lead physiologist who established the CPET test as a part of the trial and completed testing and analysis.

Bayfield, K., et al., *P207 Reliability Of Measurements Using Innocor Breath By Breath Analyser During A Maximal Exercise Test In Cystic Fibrosis Patients*. Thorax, 2014. **69**(Suppl 2): p. A167-A168.

This work was presented in poster format at the British Thoracic Society 2014 Winter meeting.

Armstrong, D., **Bayfield, K.**, et al., *P209 Standardisation Of Lung Clearance Index In A Multicentre Clinical Trial*. Thorax, 2014. **69**(Suppl 2): p. A168-A169.

This work was presented in an e poster session at the British Thoracic Society 2014 Winter meeting. I was the lead physiologist who performed, gathered and completed MBW/LCI analysis throughout the trial.

0.7 List of Figures

Figure Number	Title	Page
1.1	Diagram of an epithelial cell with the possible Cystic Fibrosis Transmembrane Conductance Regulator protein class mutations	26
1.2	The clinical progression of CF lung disease at the epithelial cells	28
1.3	An illustration of the airway tree showing the conducting and peripheral airways	34
1.4	Diagrammatic representation of lung volumes	35
1.5	Comparison of FEV ₁ and LCI Z scores from Kraemer <i>et al.</i> 2005 [1].	42
1.6	A nitrogen expirogram with calculation of the phase III slope	43
2.1	Innocor device set up indicating the patient interface and gas delivery method	73
2.2	Photograph of the modified Innocor device set up	74
2.3	Screenshot of the Simple Washout software used for analysis of MBW traces from the modified Innocor device	76
2.4	The patient interface used for the Exhalyzer D system	77
2.5	Photograph of the Exhalyzer D device set up	81
2.6	A screenshot of the sampling page from the Exhalyzer D system	82
2.7	A screenshot of the analysis page from the Exhalyzer D	83
2.8	Example traces of signal misalignment, cough, leak or large breaths on the Exhalyzer D system.	84
2.9	The set up of simultaneous gas washout using the modified Innocor device, the Exhalyzer D and the respiratory mass spectrometer	86
2.10	Screenshot of the online data monitoring function on the Respiratory Mass Spectrometer device	88
2.11	Screenshot from simple washout software loaded with the raw SF ₆ data from the Respiratory Mass Spectrometer	89
2.12	Screenshot from simple washout software loaded with the raw N ₂ data from the Respiratory Mass Spectrometer	90
2.13	The flowmeter of the Exhalyzer D (ExD/N ₂ , black box) and modified Innocor (Inn/SF ₆ , white box) were attached in series with airtight plastic tubing for simultaneous washout	99
2.14	Deadspace for the simultaneous washout of SF ₆ and N ₂ using 100% O ₂	100
3.1	LCI and LCI CV% results from novice patients	103
3.2	Bland Altman plot of CF LCI results from 3 or 2 tests using the Innocor SF ₆ device and the Exhalyzer D device	104
3.3	Screenshot of the “Z score LCI calculator” created from 42 HC participants	111
3.4	Comparison of Z score normal distribution data	112

4.1	Diagrammatic example of sources of tissue N ₂ that may contribute to and extend the washout signal in N ₂ MBW.	122
4.2	Diagrammatic representation of the potential partial washin of SF ₆ to only the larger communicating lung units.	123
4.3	Process diagram of the three main aims and subsections for the N ₂ and SF ₆ section of this thesis.	125
4.4	Bland Altman plot of the LCI difference between ExD/N ₂ and Inn/SF ₆ in HC and CF patients	128
4.5	Bland Altman FRC difference between Inn/SF ₆ and ExD/N ₂ in HC and CF	129
4.6	CF patients, plot of LCI _{2.5} vs. FEV ₁ % and LCI ₅ vs. FEV ₁ %. Results from the Inn/SF ₆ and the ExD/N ₂ devices.	130
4.7	Plot of LCI _{2.5} vs. FEV ₁ Z scores from the Inn/SF ₆ and the ExD/N ₂ .	131
4.8	Individual illustration of the discrepancy seen between Inn/SF ₆ LCI ₅ to LCI _{2.5} and ExD/N ₂ LCI ₅ to LCI _{2.5}	132
4.9	Linear regression of FEV ₁ % and LCI from the Exhalyzer D and modified Innocor devices in 44 CF patients, split into severity groups	135
4.10	Linear regression between FEV ₁ GLI Z scores and LCI (HC data from this thesis) from the Exhalyzer D and modified Innocor device. 44 CF patients, split into severity groups.	136
4.11	Linear regression of FEV ₁ % and FEV ₁ Z score against LCI from the Exhalyzer D device in 84 CF patients, split by severity groups	137
4.12	HC and CF FRC results obtained from the Innocor SF ₆ device, Exhalyzer D N ₂ device and the FRC from body plethysmography	140
4.13	Diagrammatic representation of the washin of SF ₆ followed by simultaneous washout of SF ₆ and N ₂ .	141
4.14	Graphs of individual HC and CF simultaneous washouts of SF ₆ and N ₂ on a normalised and log scaled graph.	143
4.15	Washout breath number for 7 HC subjects completing 4% SF ₆ washouts using room air or 100% O ₂ and N ₂ washout with 100% O ₂ or N ₂ and SF ₆ washout with 100% O ₂	145
4.16	Breath number differences during a 1% simultaneous washout of SF ₆ and N ₂ was completed using the Innocor, Exhalyzer D and the respiratory mass spectrometer	147
4.17	Graph of a healthy control simultaneous washout of SF ₆ and N ₂ on a normalised log scaled graph on the RMS, Innocor and Exhalyzer D	148
4.18	Summary of the different sections in SF ₆ and N ₂ chapter.	170
5.1	Demonstration of the lung dose using the blinded AeroEclipse II breath actuated nebuliser.	182
5.2	Consort diagram of patient progression throughout the trial. 116 patients formed the "Per protocol" group receiving at least 9 doses throughout the trial.	184
5.3	Graphic A shows the FEV ₁ % over the time course of the trial (relative change). Graphic B and C show the individual patient responses pre and post trial.	186

5.4	Time course of the primary outcome response stratified by baseline FEV ₁ % at trial entry.	187
5.5	Time course of the response of LCI (absolute change) to either Placebo or Active treatment.	188
5.6	Graph of FEV ₁ % and LCI from the whole group at baseline, split by severity.	193
5.7	A graph to show pre-dosing correlation of FEV ₁ % and LCI in active and placebo groups.	194
5.8	Pre dosing correlations of FEV ₁ % and LCI with other outcome measures, MEF ₂₅₋₇₅ , small mucus plugs, large mucus plugs and gas trapping	195
5.9	Time course of the LCI response stratified by baseline FEV ₁ % at trial entry	197
5.10	Comparison of LCI and FEV ₁ % values for Active and Placebo groups at Dose 4.	198
5.11	Linear regression analysis of delta changes in FEV ₁ % and LCI from pre dosing to dose 4 visits in both treatment groups	199
5.12	Comparison of Active and Placebo dose 4 LCI results in those patients with severe or mild disease	200
5.13	Within test LCI coefficient of variation (CV%) for Active and Placebo groups at each visit	201
5.14	Graph to show the proportion of people and the number of days until their first exacerbation.	203
5.15	Active and Placebo LCI waterfall plot showing the percentage change of LCI over the trial.	204
5.16	Post-dosing correlation of FEV ₁ % and LCI in both active and placebo groups.	205
5.17	Percentage delta changes in FEV ₁ % vs. absolute delta changes in LCI.	207
5.18	Starting FEV ₁ % vs. FEV ₁ % delta change.	208
5.19	Starting LCI vs. LCI delta change.	208
5.20	Comparison of both delta changes in FEV ₁ % and LCI. Delta changes of FEV ₁ % and LCI across the trial were ranked from the most improved to the most deterioration. Each individual subject was then linked and movement between quartiles was recorded.	209
5.21	Linear regression analysis of delta FEV ₁ % and LCI with MEF ₂₅₋₇₅ , small mucus plugs, large mucus plugs and gas trapping.	211
5.22	Patients with quality of life scores above the minimal clinically important difference were selected and the positive and negative changes in FEV ₁ % and LCI plotted.	213
5.23	ROC curves of CFQR QoL respiratory domain vs FEV ₁ % and LCI.	214
5.24	Graph to show the average FEV ₁ % from run in, multi-dose and follow up data in both active and placebo patients.	218
5.25	Average LCI from run in and multi-dose in both active and placebo patients.	219

6.1	FEV ₁ % and LCI changes from visit 1 to visit 2 after an intervention.	238
6.2	Delta changes for FEV ₁ % and LCI in 32 patients who had a treatment intervention.	238
6.3	FEV ₁ % and LCI results from G551D patients who started ivacaftor treatment.	239
6.4	FEV ₁ % and LCI at time point 1 and time point 2 from those with more severe and milder disease patients.	240
6.5	FEV ₁ %, LCI and CFQR QoL results in 16 patients with a treatment intervention.	242
6.6	Percentage of people with improvements or deterioration in LCI and FEV ₁ in those patients with improved CFQR QoL above the MCID	242
6.7	Three individual people monitored for occurrences of ABPA.	244

0.8 List of Tables

Table Number	Title	Page
1.1	Key interventional studies where FEV ₁ treatment changes are seen after an intervention	39
1.2	LCI treatment changes after an intervention	53
2.1	European Respiratory Society/ American Thoracic Society, a summary of spirometry guidelines	64
2.2	European Respiratory Society/ American Thoracic Society, a summary of spirometry guidelines	65
2.3	Gas properties of the gases utilised in multiple breath washout test	70
2.4	Table of device specifications utilised within this thesis	93
3.1	HC circadian rhythm data	107
3.2	CF circadian rhythm data	107
4.1	Key papers that complete LCI measurements on both healthy control and CF patients in a variety of ages, different devices, gases and utilising different software algorithms	118
4.2	Subject demographics for CF and HC individuals who completed at least two MBW tests on the Exhalyzer D and modified Innocor device on the same day	127
4.3	LCI _{2.5} and LCI ₅ results for CF and HC participants	128
4.4	Innocor and Exhalyzer D differences in FRC, CEV, Respiratory Rate and tidal volume in both healthy control and cystic fibrosis patients	129
4.5	CF Patient demographics for groups split by severity <60%, 60-80% and >80% FEV ₁ predicted.	134
4.6	Demographic data from HC and CF patients where FRC results from Innocor, Exhalyzer D and body plethysmography	139
4.7	Demographic data of subjects completing simultaneous washout of SF ₆ and N ₂ .	142
4.8	Demographic data of HC subjects who completed the 1% SF ₆ and N ₂ washout.	146
5.1	Inclusion and Exclusion criteria for the cystic fibrosis gene therapy multi-dose trial (MDT)	176
5.2	Outcome measures in the MDT	178
5.3	Acceptability criteria for LCI inclusion.	179
5.4	Summary of study assessments	181
5.5	Baseline characteristics of the 116 participants in the MDT	185
5.6	Absolute and standardised treatment effects of the major primary and secondary outcomes in the trial (FEV ₁ %, LCI, MEF ₂₅₋₇₅ , CT parameters (Small mucus plugs, Large mucus plugs, gas trapping) and the respiratory domain from the CFQR QoL.	189

5.7	LCI and FEV ₁ % coefficient of variation results pre and post Dose 4.	202
5.8	Post trial linear regression slopes for FEV ₁ % and LCI vs. other secondary outcome measures.	206
5.9	Table of Demographic data from 34 people who were in the run in study and went on to complete >9 doses of gene therapy or placebo doses during the multi-dose trial.	216
6.1	Recruitment numbers per treatment intervention and patient breakdown of the three main groups in the study.	236
6.2	Patient demographic data for each treatment intervention group.	237
6.3	FEV ₁ and LCI treatment responses to individual treatment interventions.	239

0.9 Table of Contents

0	Title Page.....	0
0.1	Declaration of copyright.....	1
0.2	Declaration of originality.....	1
0.3	Abstract.....	2
0.4	Acknowledgements.....	3
0.5	Abbreviations.....	4
0.6	Publications and Presentations arising.....	8
0.7	List of Figures.....	11
0.8	List of Tables.....	15
0.9	Table of Contents.....	17

1	Chapter 1 – Introduction.....	25
1.1	Cystic Fibrosis (CF).....	25
1.1.1	Genetics.....	25
1.1.2	Disease Pathophysiology.....	26
1.1.2.1	Mucus.....	27
1.1.3	Pulmonary exacerbations, Infection and Inflammation.....	27
1.1.4	Diagnosis and Disease Manifestation.....	28
1.2	Conventional and Novel therapy.....	29
1.2.1	Conventional Therapy.....	29
1.2.2	Novel Small Molecule Therapy.....	30
1.2.3	Monitoring disease in clinic.....	31
1.3	Lung physiology and physiological tests.....	32
1.3.1	Spirometry.....	33
1.3.2	Plethysmography.....	35
1.4	Computed Tomography.....	36
1.5	Patient Reported outcomes.....	36
1.6	Monitoring mild lung disease.....	37
1.7	Concept of a minimal clinically important difference.....	37
1.8	Multiple breath washout – Lung clearance index.....	41
1.8.1	Ventilation inhomogeneity.....	42
1.8.2	Multiple breath washout testing.....	44

1.8.2.1	Determining the end of the test.....	45
1.8.2.2	Devices.....	45
1.8.3	Parameters of the multiple breath washout test.....	45
1.8.3.1	Lung clearance index.....	45
1.8.3.1.1	What is LCI measuring?	46
1.8.3.2	Further analysis of the MBW trace.....	47
1.9	Clinimetric properties of Lung Clearance Index.....	48
1.9.1	Feasibility.....	49
1.9.2	Validity.....	49
1.9.3	Reliability.....	50
1.9.4	LCI and relationship with other outcomes.....	50
1.9.5	LCI treatment responses.....	51
1.10	Standardisation of MBW.....	57
1.10.1	Guidelines.....	57
1.10.2	SF ₆ and N ₂	58
1.10.2.1	SF ₆	58
1.10.2.2	N ₂	58
1.10.2.3	SF ₆ and N ₂ MBW comparison.....	59
2	Chapter 2 - Methodology.....	61
2.1	Study Subjects.....	61
2.2	Demographic Data.....	62
2.3	Outcome measures.....	62
2.3.1	Spirometry.....	62
2.3.2	Plethysmography.....	65
2.3.3	Computed Tomography.....	66
2.3.4	Quality of Life Questionnaire.....	66
2.4	Multiple Breath Washout.....	67
2.4.1	Breathing pattern.....	67
2.4.2	Deadspace.....	68
2.4.3	Gases.....	68
2.4.3.1	Tracer Gas.....	69
2.4.3.2	Washout Gas.....	69
2.4.4	Innocor SF ₆ Device.....	70

2.4.4.1	Device.....	70
2.4.4.2	Calibration.....	72
2.4.4.3	Testing.....	74
2.4.4.4	Analysis.....	75
2.4.5	Exhalyzer D N ₂ Device.....	76
2.4.5.1	Device.....	77
2.4.5.2	Calibration.....	78
2.4.5.3	Testing.....	80
2.4.5.4	Analysis.....	81
2.4.6	Respiratory Mass Spectrometry.....	84
2.4.6.1	Device.....	85
2.4.6.2	Calibration.....	85
2.4.6.3	Testing.....	85
2.4.6.4	Analysis.....	86
2.4.7	Device Specifications.....	91
2.4.8	Simultaneous Testing.....	98
2.4.9	Statistics.....	101

3 Chapter 3 – Method Development.....102

3.1	Introduction.....	102
3.2	Novice patients.....	102
3.2.1	Hypothesis, Aims and Objectives.....	102
3.2.2	Methods.....	102
3.2.3	Results.....	102
3.2.4	Summary.....	103
3.3	Two or Three Multiple Breath Washout Tests.....	103
3.3.1	Hypothesis, Aims and Objectives.....	104
3.3.2	Methods.....	104
3.3.3	Results.....	104
3.3.4	Summary.....	105
3.4	Circadian Rhythm and Lung Clearance Index.....	105
3.4.1	Hypothesis, Aims and Objectives.....	106
3.4.2	Methods.....	106
3.4.3	Results.....	106

3.4.4	Summary.....	107
3.5	Lung Clearance Index Z scores.....	109
3.5.1	Hypothesis, Aims and Objectives.....	110
3.5.2	Methods.....	110
3.5.3	Results.....	111
3.5.4	Summary.....	113

4 Chapter 4 - Defining and explaining differences in MBW

	observed with different techniques.....	116
4.1	Introduction.....	116
4.2	Hypotheses, Aims and Objectives.....	124
4.3	Methods.....	125
4.4	Results.....	126
4.4.1	Compare N ₂ and SF ₆ (Aim 1)	126
4.4.2	Examine whether the differences between devices are impacted by disease severity (Aim 2).....	133
4.4.3	Understanding LCI differences between gases: (Aim 3)	138
4.4.3.1	Does FRC cause the difference between devices?.....	138
4.4.3.2	Do individual patient factors between tests contribute to the difference? Simultaneous washout with an extended washout period.....	141
4.4.3.3	Respiratory Mass Spectrometry.....	144
4.4.3.3.1	Do tracer gases washout differently? Dose breathing 100% O ₂ impact the washout?.....	144
4.4.3.3.2	Does using the respiratory mass spectrometer to measure both gases eliminate equipment differences?	145
4.5	Discussion.....	148
4.5.1	Comparison of N ₂ and SF ₆ LCI results.....	148
4.5.1.1	LCI _{2.5} and LCI ₅	150
4.5.2	LCI results and disease severity.....	150
4.5.3	Understanding LCI differences between gases.....	154
4.5.3.1	FRC.....	154
4.5.3.2	Differences related to individuals' breathing patterns?.....	155
4.5.3.2.1	Simultaneous washout.....	156

4.5.3.3 Tissue N ₂	156
4.5.3.3.1 Limitations with simultaneous washout testing.....	159
4.5.3.4 Respiratory Mass Spectrometry – Differences in gases.....	161
4.5.3.4.1 Molecular properties of gases.....	161
4.5.3.4.2 Washout gas.....	163
4.5.3.4.3 Viscosity.....	164
4.5.3.5 Respiratory Mass Spectrometry – Do equipment differences explain N ₂ and SF ₆ differences?.....	165
4.5.3.5.1 RMS Limitations.....	166
4.5.4 Chapter conclusions.....	167
4.5.5 Future Studies for the understanding of N ₂ and SF ₆ differences.....	169

5 Chapter 5 - Exploring Lung Clearance Index as an outcome measure in the Cystic Fibrosis Gene Therapy Multi-dose Trial (MDT)

5.1 Introduction.....	171
5.1.1 Gene Therapy and the cystic fibrosis gene therapy consortium.....	171
5.1.1.1 Gene delivery agents.....	172
5.1.1.2 GL67A.....	172
5.1.1.3 pGM169.....	173
5.1.1.4 Single Dose trial.....	173
5.1.1.5 Physiological Outcome Measures.....	174
5.1.2 The role of the author on the MDT.....	174
5.2 Hypotheses, Aims and Objectives.....	175
5.3 Methods.....	175
5.3.1 Study Design and Study subjects.....	175
5.3.2 Randomisation.....	176
5.3.3 Trial Triage.....	177
5.3.4 Outcome Measures.....	177
5.3.4.1 Primary outcome – FEV ₁	177
5.3.4.2 Secondary outcomes.....	178
5.3.4.2.1 Lung Clearance Index.....	178
5.3.4.2.2 Chest CT Scan.....	179

5.3.4.2.3	Frequency of respiratory exacerbations.....	179
5.3.4.2.4	Quality of Life Questionnaire.....	179
5.3.4.2.5	Confirmation of gene transfer.....	180
5.3.5	Dose administration.....	180
5.3.6	Data Analysis and Statistics.....	182
5.4	Results.....	183
5.4.1	Patient groups and Demographics.....	183
5.4.2	Primary Outcome.....	185
5.4.3	Major Secondary Outcomes.....	187
5.5	Further analysis for this thesis.....	190
5.5.1	Research questions.....	190
5.6	Hypotheses, Aims and Objectives.....	190
5.7	Analysis methods for this thesis.....	191
5.8	Results.....	192
5.8.1	Baseline Characteristics from an LCI perspective.....	192
5.8.2	Pre dose correlations.....	193
5.8.3	LCI change over the trial.....	196
5.8.3.1	LCI and severity.....	196
5.8.3.2	LCI results at an earlier time point.....	197
5.8.3.2.1	Was the difference at Dose 4 more pronounced in one severity group?.....	199
5.8.3.2.2	Was the difference due to stability at the beginning of the trial?.....	200
5.8.3.2.2.1	Within test variability.....	200
5.8.3.2.2.2	Between test variability.....	201
5.8.3.2.3	Was there a difference in LCI at an earlier time point because patients had not yet experienced a pulmonary exacerbation?.....	202
5.8.3.3	Delta changes in LCI.....	203
5.8.3.3.1	How did changes in LCI and FEV ₁ relate?.....	204
5.8.3.3.1.1	Post Trial Correlations.....	204
5.8.3.3.1.2	Did FEV ₁ % and change similarly across the trial?	206
5.8.3.3.1.3	Did a subjects starting lung function affect their delta change across the trial?.....	207
5.8.4	Correlation with other outcomes.....	209
5.8.4.1	Physiological outcome measures.....	210
5.8.4.2	CFQR QoL.....	212

5.8.5	Run in results, did the MDT change the trajectory of FEV ₁ and LCI?.....	214
5.9	Discussion.....	220
5.9.1	Patient recruitment, retention and safety.....	220
5.9.2	Demographics.....	221
5.9.3	Main trial outcomes (primary and secondary)	222
5.9.4	What can we learn from the MDT about LCI as a clinical trial outcome?.....	223
5.9.4.1	LCI and FEV ₁ mismatch.....	224
5.9.4.2	LCI is unresponsive to Gene Therapy treatment.....	225
5.9.4.3	Sample size; is LCI underpowered?	225
5.9.4.4	Severity.....	226
5.9.4.5	Variability of LCI response and natural disease progression.....	227
5.9.4.6	LCI results at an earlier time point.....	229
5.9.5	Patient reported outcomes – CFQR QoL.....	230
5.9.6	Did the MDT change the trajectory of FEV ₁ or LCI?.....	231
5.10	Summary.....	232

6 Chapter 6 - Can we use the response in LCI to treatments to help define a clinically meaningful change?.....233

6.1	Introduction.....	233
6.2	Hypotheses, Aims and Objectives.....	233
6.3	Methods.....	234
6.3.1	Study subjects and treatment changes.....	234
6.3.2	Study design.....	234
6.3.3	Other outcome measures.....	235
6.3.4	Statistics.....	235
6.4	Results.....	235
6.4.1	Demographics.....	235
6.4.2	Anticipated treatment group.....	237
6.4.2.1	FEV ₁ and LCI results.....	237
6.4.2.2	Within group changes.....	238
6.4.3	Severity.....	239
6.4.4	Stability.....	240
6.4.5	Results assessed alongside CFQR QoL.....	241

6.4.6	Dry powder inhalers.....	243
6.4.7	ABPA.....	243
6.5	Discussion.....	245
6.5.1	Did the duration of the intervention impact the LCI result?.....	245
6.5.2	Is LCI ‘too sensitive’ to demonstrate an improvement after treatment?.....	246
6.5.3	Do drugs with different mechanisms of action differentially affect LCI?.....	247
6.5.4	Does severity impact the LCI response to treatment?.....	248
6.5.5	Does clinical stability influence the LCI response?.....	249
6.5.6	CFQR QoL results after treatment.....	250
6.5.7	LCI and FEV ₁ after powdered inhalers.....	252
6.5.8	Does LCI help to assess ABPA treatment?	253
6.6	Summary of treatment intervention section.....	255
7	Chapter 7 – Overall Summary and Conclusions.....	256
7.1	Overarching conclusions.....	256
7.2	The future for Lung Clearance Index as an outcome measure.....	257
7.2.1	SF ₆ and N ₂	257
7.2.2	Interventional studies required.....	257
7.2.3	Identifying the optimal patient population for added benefit of LCI.....	258
7.2.4	Understanding the regional ventilation inhomogeneity of LCI.....	258
7.2.5	Standardisation.....	259
8	References.....	260
9	Appendices.....	274

1 Chapter 1 – Introduction

1.1 Cystic Fibrosis (CF)

CF is a lethal autosomal recessive condition affecting >10,000 people in the UK alone (UK CF Registry data report 2015 [2]). It is the most common genetic disease in the Caucasian population with the prevalence varying throughout Europe [3]. CF is a progressive disease that has notoriously been regarded as a paediatric disease, yet the median survival in CF in the UK is now 35 years showing that strategies for monitoring and treating disease have improved and are changing [4].

1.1.1 Genetics

CF is caused by a mutation on the long arm of chromosome 7 causing incorrect gene sequencing and misfolding of the Cystic Fibrosis Transmembrane Conductance Regulator (CFTR) protein. There are different classes of mutations based on the amount of protein produced and the location of the abnormality, Figure 1.1 shows the 6 classes of mutations [5]. The most common mutation is F508del (class II) where the protein does not fold properly and therefore does not escape the endoplasmic reticulum to reach the cell surface for protein function. Over 2000 different mutation variants have been identified but not all of them are disease causing.

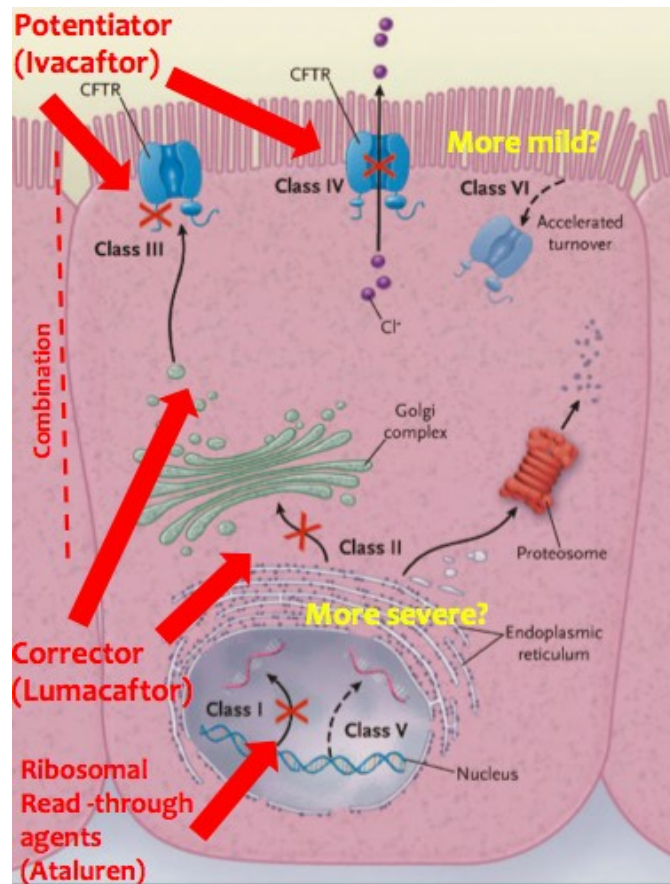


Figure 1.1 Diagram of an epithelial cell with the possible Cystic Fibrosis Transmembrane Conductance Regulator protein class mutations [5]. Labels show the potential small molecule mechanisms of action. Reproduced with permission from Rowe *et al*, 2005. Copyright Massachusetts Medical Society.

1.1.2 Disease Pathophysiology

The CFTR gene codes for a cyclic adenosine monophosphate (cAMP) regulated channel on epithelial cell surface of a number tissues, the airway epithelial cells, sweat glands, pancreatic and biliary ducts, intestinal mucosa and the male vas deferens [6]. It is responsible for the efflux of chloride and bicarbonate ions from the cells, it also plays a role in the transport of sodium, across the cell membrane with inhibition of epithelial sodium channel (ENaC) [7]. In CF sweat contains 3-5 fold the amount of sodium chloride compared to healthy individuals and this was the cornerstone of CF identification by Dorothy Anderson in 1938 [8]. The increase in sodium chloride shows the abnormal CFTR protein activity, however it is the secondary morphological changes in the lungs and pancreas that seem to show inherent disturbances in fluid transport that make CF disease fatal [9].

This flux of sodium from a high concentration to a low concentration across the epithelia mean that the airway surface liquid is depleted [10]. This is deemed the low volume hypothesis of lung disease

[5]. This hypothesis suggests that water is drawn across the membrane of the cells due to the salt concentration gradient and because of this the cilia on the epithelial cells collapse and therefore the cilia cannot beat normally and clear mucus at an optimal level.

1.1.2.1 Mucus

A protective mucus lining is released on cell surfaces of the lungs and intestines with the benefit of trapping and eliminating inhaled particles as well as preventing desiccation of airway surfaces [11]. As described, when the CF airways are dehydrated then cilia on the cell surface are not able to beat and promote normal mucociliary clearance so there is mucus build up within the lungs. The mucus also becomes dehydrated due to the salt concentration gradient. Bicarbonate is known to be important in the normal expansion of the glycoproteins (mucins) that make up mucus [12]. Bicarbonate transport is also defective with the CFTR mutation and it has been proposed that without bicarbonate, mucus is thicker, poorly solubilised and less transportable within the lung [12]. This thick, sticky, dehydrated mucus within the warm environment of the body tissues is ideal for bacterial pathogens to grow and colonise.

1.1.3 Pulmonary exacerbations, Infection and Inflammation

People with CF have recurrent exacerbations likely due to the increased mucus and the potential for bacterial infection. Therefore increased symptoms of chronic CF disease i.e. cough and sputum production are likely. Pulmonary exacerbations are best characterised as a constellation of acute changes in clinical status and pulmonary function away from a patients normal stability [13]. Early infection at infancy has been shown to impact lung function in childhood [14], showing that early intervention and prevention is essential to maintain respiratory health status. A questionnaire has been developed including information about symptoms, presence, absence or severity of an exacerbation, physical examination and pulmonary function tests [13] but ultimately the classification of a pulmonary exacerbation is based on clinical judgement and decision to treat [15] so remains subjective. Lung function loss tends to occur during or after a pulmonary exacerbation [16, 17], with as many as 25% of patients not returning to 90% of their baseline 3 months after treatment [18]. Once chronic infection is established then a significantly lower baseline and faster deterioration in lung function is seen in those who are colonised vs. those without infection [19]. The most common chronic infection is *pseudomonas aeruginosa*, peaking at age 32-35 years, but other pathogens have also shown negative effects on lung function [2]. These persistent infections cause hypersensitive inflammatory and neutrophilic responses and the chronic inflammation causes pathological manifestations in the airways and ultimately, bronchiectasis, irreversible airway wall

thickening and dilatation increases [8, 20]. This CF disease pathogenesis is summarised in Figure 1.2, showing a diagram of chronic disease progression at an epithelial level [21].

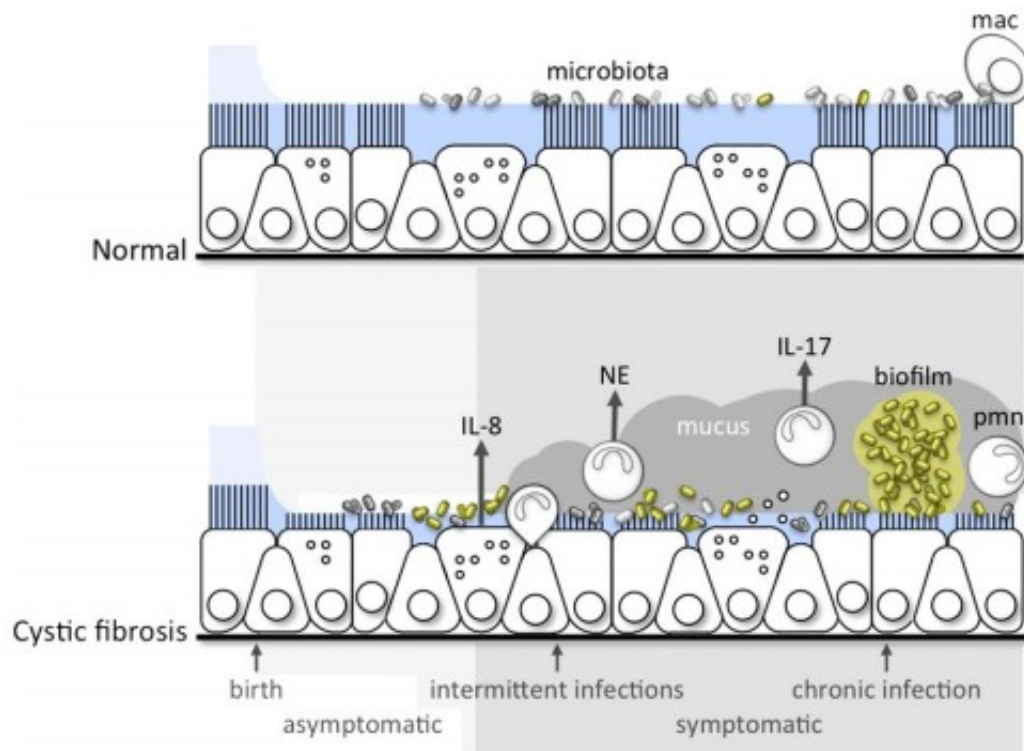


Figure 1.2 The clinical progression of CF lung disease at the epithelial cells; lower air surface liquid volume, disruption of mucociliary clearance leading to intermittent then chronic infection with inflammation. Mac= macrophage, NE = neutrophil elastase, pmn = neutrophil. Reprinted with permission of the American Thoracic Society. Copyright © 2017 American Thoracic Society. *Annals of the American Thoracic Society* is an official journal of the American Thoracic Society [21].

1.1.4 Diagnosis and disease manifestation

Newborn screening for CF has been introduced fairly recently and has improved early disease by many new-borns no longer presenting with early characteristic symptoms i.e. failure to thrive or persistent cough [22]. Sweat chloride testing is the gold standard to diagnose CFTR function [23]. Sweat inducing pilocarpine iontophoresis is used to stimulate CFTR function in the sweat gland. If CFTR function is abnormal there will be a positive sweat chloride result (>60 mmol/L). Practical guidelines can be found here: http://www.acb.org.uk/whatwedo/science/best_practice/acb-developed-guidelines. Not all sweat chloride results are conclusive and repeat testing is performed, if the result is inconclusive again then other tests such as Genetics testing are completed (if not completed in parallel) to determine each individual variant and the potential associated clinical

phenotypes. The CFTR2 website aims to provide clinical information about different genotypes and prevalence of clinical signs and symptoms (<https://www.cftr2.org/resources>) it has helped in the differential diagnosis of individuals with non-classical CF disease. Individuals may also have nasal potential difference (NPD) tests completed (CFTR function stimulated by salt solutions on the nasal epithelium) or intestinal current measurements to test the CFTR function on rectal mucosa biopsy samples. The Royal Brompton clinical guidelines on CF diagnosis can be found here:

<http://www.rbht.nhs.uk/healthprofessionals/clinical-departments/cystic-fibrosis/clinical-cf-guidelines-care-of-children/making-the-diagnosis/>.

CF is a multi-system disease, with many patients having pancreatic insufficiency, Sino-pulmonary disease, CF related liver disease, low bone mineral density and male infertility however the predominant manifestation of the defect is in the airway epithelial of the lungs [8]. Therefore the lungs are the predominant downstream target of treatment as respiratory failure is the main cause of death in CF [24].

Structural lung changes occur from a young age in CF (0-4months) with evidence of bronchial wall inflammation as detected by high-resolution computed tomography (CT) [25]. These abnormalities have led groups to suggest that the optimal treatment window is before 12 months, or before structural, functional or biochemical changes can be detected [21]. Currently, treatment concentrates on prophylaxis to slow the rate of decline in the clinical condition and many with people with CF now have mild manifestations of the disease due to diligent and early specialist care from multidisciplinary teams [8]. In fact, adults with CF now outnumber children [24] and so the slow progression of the disease [26] needs to be closely monitored with sensitive measures to continue to improve outcomes in children and adults [27].

1.2 Conventional and Novel therapy

1.2.1 Conventional Therapy

Whilst CF is a multi-organ disease and optimising nutrition has been shown to improve outcomes in CF [28], the main treatment regimes target the lungs. The mainstays of therapy include regular mucus clearance with physiotherapy techniques, mucolytics (Dornase alpha (DNase) and hypertonic saline) and antibiotics, via oral, inhaled and intravenous routes. The main aims of these therapies are to maintain clear, non-infected airways and prevent/delay irreversible scarring and bronchiectasis (downstream manifestations of the disease). Treatment is individualised depending on the disease stage, nature of bacterial/ fungal infection, the occurrence of exacerbations and what colonisations of bacteria or fungi (and to some extent patient preference and ability to adhere). The standard treatments of an exacerbation are intravenous antibiotics and mucolytics; although once a patient is

chronically infected with pseudomonas, inhaled antibiotics are part of a patient's routine medications. Conventional therapies i.e. mucolytics (Dornase alpha or hypertonic saline) and antibiotics (Tobramycin, Colomycin or Astreonam) generally result in a modest improvement of 3-5% in lung function [8].

DNase is an enzyme that breaks down extracellular DNA; therefore the viscoelasticity of lung mucus secretions can be reduced [29] and mucociliary clearance increased. DNase is started at age 6 in CF patients at the Royal Brompton Hospital (RBH), for the optimisation of respiratory health (<http://www.rbht.nhs.uk/healthprofessionals/clinical-departments/cystic-fibrosis/clinical-cf-guidelines-care-of-children/>). Hypertonic saline and Mannitol aim to re-hydrate the airways and restore mucociliary clearance [30, 31].

Despite these early and preventative therapies, there is build up of thick sticky mucus in CF airways and with that comes increased bacterial infection. Different antibiotics are used to target specific bacteria either intravenously (IV's) or orally but the regimes are not necessarily standardised between or within sites [32]. There is also increased interest in whether antibiotics should be delivered via traditional nebulised or a relatively new delivery method, powdered inhalers. It has been shown that powdered inhalers reduce treatment time and are tolerable for CF patients [20] as well as showing improvements in absolute change in forced expiratory volume in one second percentage predicted (FEV₁%) and reduction in pseudomonas sputum density using the anti pseudomonal agent colistimethate sodium [33]. As of yet there is little evidence on the impact of switching from one to the other or if one is advantageous over another.

1.2.2 Novel Small molecule Therapy

At the forefront of therapy are small molecule treatments that target specific CFTR mutations (Figure 1.2) rather than downstream manifestations of the disease [34]. The potentiator, Ivacaftor (trade name Kalydeco) targets CFTR Class III mutations and increases the time that the CFTR channels on the cell surface membrane remain open [35]. The exact mechanism of action is not completely understood but it is thought that the CFTR gating channels are opened by further phosphorylation of the CFTR protein thereby increasing its function [36]. Results have shown remarkable improvements in lung function, a relative change of FEV₁ by 17.2% compared to placebo after 24 weeks of treatment and the effect of ivacaftor was seen after just 15 days [35]. CFTR targeted therapies are expected to have the highest potential to improve patient outcomes as these results are beyond the usual improvements from routine treatments [37]. Improvements have also been shown with this targeted drug to those with preserved lung function by using lung clearance index [38] for which this thesis concentrates. The systemic treatment showed sustained

improvements in time to pulmonary exacerbations, increased weight, positive patient reported respiratory symptoms and decreased sweat chloride results [35] showing that targeted treatment provides clear proof of concept for these approaches [34]. It is important to note that only 4% of the CF population have CFTR mutations that would respond to ivacaftor however [39] so the treatment paradigm as it stands and improvements to therapies for all aspects of CF disease are still required [37].

Other small molecules include CFTR correctors that decrease the misfolding of the protein and increase the trafficking of the protein to the cell surface [40]. Correctors are used in combination therapies with a potentiator (Lumacaftor and Ivacaftor (Trade name Orkambi) [40-42]). But so far, these agents have not shown as much clinical utility as ivacaftor alone, potentially because the targeted classes and genotypes are considered to have a more severe phenotype than classes III or IV [43]. In the UK although Orkambi is licensed, the National Health Service (NHS) does not yet provide it, this is likely because the level of benefit is not great enough for the cost. The smaller response to Lumacaftor, although still in line with routine treatments when paired in combination with ivacaftor, may be due to the multistep process of both corrector and potentiator. However improvements were seen when young children with mild disease were tested [41], providing evidence that early targeted treatment may have the greatest potential to prevent organ damage [34].

There are also ribosomal read-through agents (Ataluren [44]) for class I mutations. But the group suggested that there was a lack of response when patients were also on inhaled tobramycin, as competition for the ribosomal binding site was increased [44] although a recent press release stated that a follow on trial failed to reach its primary endpoint so small molecule targeted treatment in this class mutation may be limited (<http://ir.ptcbio.com/releasedetail.cfm?ReleaseID=1015471>). Gene Therapy is also a potential treatment of the basic CFTR defect and this will be discussed in Chapter 5.

1.2.3 Monitoring disease in the clinic

At RBH children are seen at regular 3 monthly intervals by a clinic visit with the full multidisciplinary team. Dieticians monitor and advise on appropriate nutritional recommendations, bacterial cultural swabs or sputum samples are taken after consulting with the physiotherapists on appropriate mucociliary clearance techniques, spirometry values are collected to monitor lung function between visits and CF specialist nurses and doctors assess any clinical signs or symptoms of deteriorating disease and prescribe medications in light of any periods of exacerbation. Psychological examinations, blood sampling, glucose monitoring for CF related diabetes, bronchoscopies for more

thorough respiratory investigation and yearly chest imaging and lung clearance index are also completed to track and monitor disease. With this in mind, CF children can show little disease progression and or clinical presentation of the disease [45]. Whilst mild disease is obviously a good presentation for those with CF, it does present difficulties in identifying and measuring small and subtle changes in disease to be able to target treatment early and prevent further lung damage. Therefore sensitive outcome measures are required.

1.3 Lung physiology and physiological tests

The anatomical structure of the lungs consist of airway bifurcations from the trachea to increase the airway surface area to facilitate gas exchange, the most commonly used schematic is by Weibel & Ewald (1963) [46]. Figure 1.3 shows the airway tree and the 23 generations of airways; 0-16 are considered the conducting airways where gas travels by convection (convection is where gas travels from high pressure to low pressure in the case of bulk flow during inspiration or expiration) and 17-23 are considered the intra-acinar or peripheral airways where around 300 million alveoli promote gas exchange by diffusion (exchange of gases from the alveoli to the blood via diffusion and affinity gradients) [47].

Small airways narrow more distally and gas mixing efficiency is heterogeneous even in healthy lungs due to structural asymmetry [48] i.e. different lengths and cross sectional areas of branching airways [49]. There is also heterogeneity at the point where gas convection changes to diffusion “the diffusion convection front” where the distribution of gas in parallel branches may be unequal when considering a single branch point model [50]. In health, gas is generally distributed relatively evenly throughout the airways but in disease with increased mucus causing obstruction and airway wall thickening and damage due to inflammation [51] air flow and gas mixing is uneven and coined ventilation inhomogeneity (VI).

An ability to monitor the declining lung health and responses to treatment are essential aspects of an improved prognosis in CF. Surrogate markers of these disease manifestations and the impact on lung function are described below.

A surrogate outcome measure is one that can be used to predict the efficacy of an intervention, when direct measurement is not possible [52]. As it stands the main surrogate outcome for detecting lung disease progression is forced expiratory volume in one second (FEV₁) [53]. The higher the result, the more preserved the lung function. Spirometry is the gold standard of lung function testing, recognised by the European Respiratory Society (ERS) and American Thoracic Society (ATS) as a standardised test useful for diagnosis i.e. evaluation of symptoms and prognosis, as well as monitoring i.e. assess disease progression or response to intervention [54]. Risk factors for

deterioration of FEV₁ include infection with *Pseudomonas aeruginosa*, female sex, poorer nutritional status as well as high starting baseline and the presence of crackles [18, 55]. The authors suggest that knowing the rate of lung function decline is not only important for the increased clinical care of patients but also for the design and analysis of clinical trials, where FEV₁ decline should be taken into consideration [18, 56].

1.3.1 Spirometry

Spirometry is a surrogate measure of lung volumes and airway patency (Figure 1.4); resting flow volume loops or forced expiratory efforts can reveal airway obstruction, restriction, or lung hyperinflation. The test also reveals if a person is performing below what would be predicted according to their height, weight, age, sex and ethnicity [57]. The ERS and ATS have published guidelines on appropriate technique, repeatability and interpretation [54, 58] (see methodology section).

Spirometry involves the patient completing an effort dependent, forceful exhalation, and it predominantly shows results from large conducting airways (generation 1-16 of the respiratory tract - Figure 1.3). Measures of more distal airways include mid expiratory flows (MEF₂₅₋₇₅) from the spirometry flow volume loop but these have been shown to be quite variable and lack consistency [59].

Percentage predicted calculations and Z scores are used to orientate the spirometry values to what is expected for an individual in relation to a healthy control population. Z scores are considered to be more robust as they anchor the result to the spread of the normal population and deviation outside ± 2 standard deviations (SD) is considered abnormal. There are many established reference ranges, the main used historically at RBH are the European Coal and Steel Community equations for adults [60] and Rosenthal for children [61]. It has long been known that switching between reference ranges produces different predicted values, which will affect interpretation and could lead to unnecessary and even detrimental clinical treatment. For example height and age can have an impact as great as 38-40% when comparing to different reference ranges [62]. Recently the ERS launched a task force to establish an all age, multi ethnic reference range for spirometry [57]. The global prediction equations enable smoothing of data from childhood to adulthood and incorporate data from multiple countries with a large age range. Data is now transferable between sites and studies and is quantifiable and comparable across a person's lifetime. Using percentage predicted is an easy way to quantify results i.e. a drop of 20% is likely to be significant (http://www.spirxpert.com/gli_percent_predicted.html) and this is easily understandable. Z scores are a much more accurate reflection of deviation however as they encompass the normal

distribution of data and those at the flanks and outside of the upper and lower limits of normal are not subject to large changes depending on where the majority of the data falls. A more comprehensive explanation of spirometry Z scores can be found in other sources [57, 63]. With the use of Z scores, results are free from bias due to age, height, sex and ethnic group.

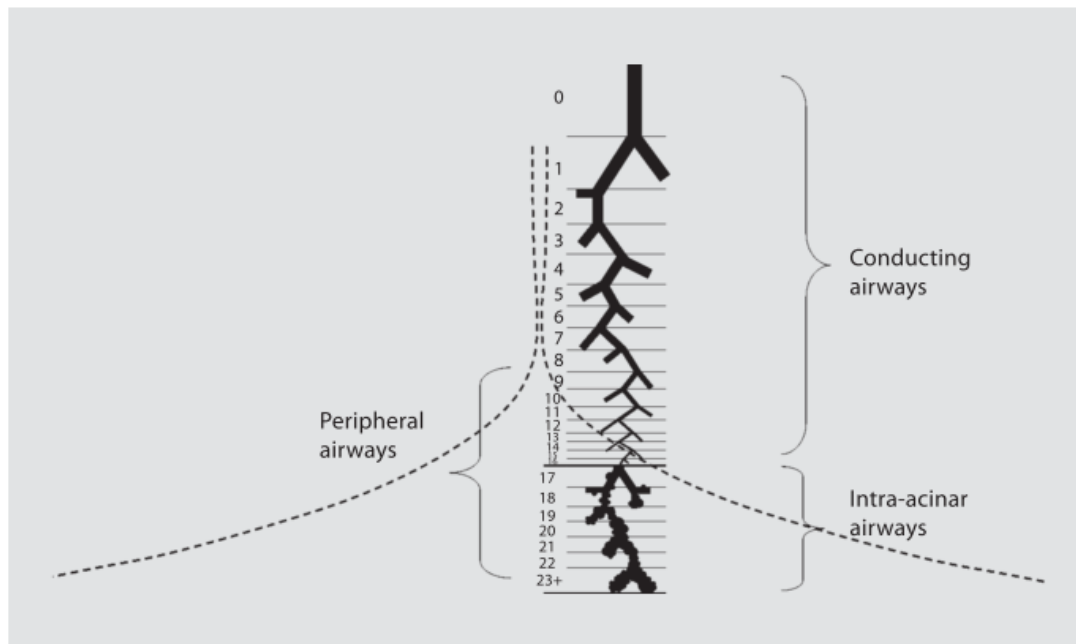


Figure 1.3 An illustration of the airway tree, branches 1-16 represent the conducting airways and 17-23+ represents the peripheral airways. The surface area of the lungs dramatically increases the further down the airways, represented by the dotted line. This Figure was reproduced with permission, Copyright © 2012 Karger Publishers, Basel, Switzerland [47].

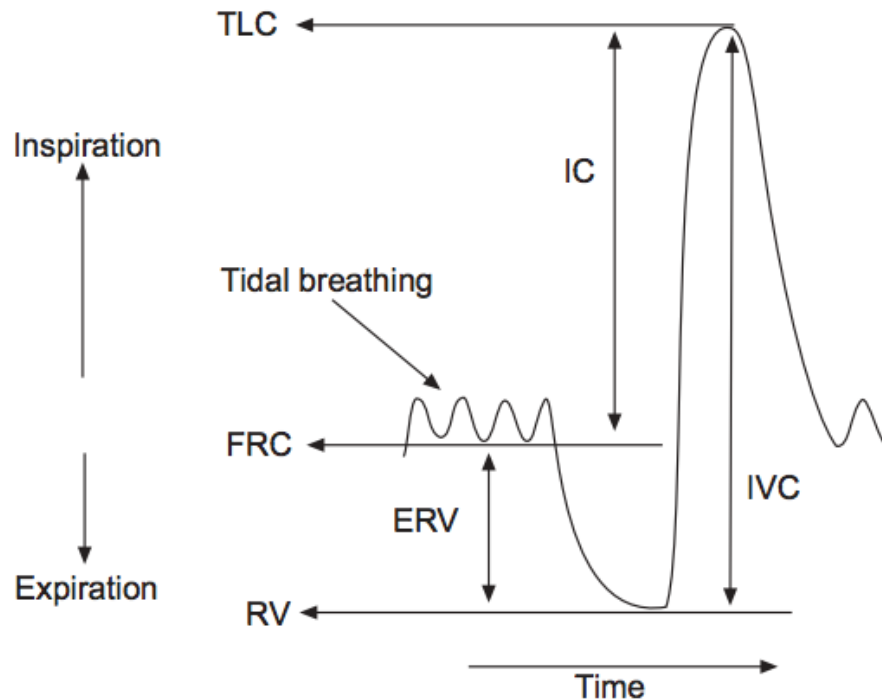


Figure 1.4 Diagrammatic representation of lung volumes, tidal breathing at rest represents approximately 500ml in a healthy adult, normal tidal breathing should return to a functional residual capacity (FRC) with each breath. A forced expiratory breath reveals the expiratory reserve volume (ERV) and the volume left in the lungs at this point is the residual volume (RV). A maximal inspiratory breath to total lung capacity (TLC) gives us the inspiratory vital capacity (IVC) and the inspiratory capacity (IC). Reproduced with permission of the European Respiratory Society ©. European Respiratory Journal Aug 2005, 26 (2) 319-338; DOI: 10.1183/09031936.05.00034805 [54].

1.3.2 Plethysmography

Plethysmography is the gold standard technique for measuring lung volumes [64]. Subjects sit inside a pressurised “body box” and complete different breathing protocols with a mouthpiece. The pressure changes at the mouth and within the box allow calculation of lung volumes (see methods section). Body plethysmography measures all of the thoracic gas therefore it could include emphysematous regions, trapped gas, pneumothorax and abdominal gas i.e. not just airways that are communicating with the bronchial tree for gas exchange [65]. Lung volumes can be exaggerated or underestimated if the panting technique is altered i.e. diaphragmatic breathing can underestimate thoracic gas volume and the test can be problematic in children [66]. Lung volumes can also be overestimated, as changes in mouth pressure may be less than alveoli pressure with airway obstruction, due to small/narrow airways, trapped gas and increased compliance of the lung. In other words there may be a delay with high airways resistance so equilibrium of alveoli and

mouth pressure may not be reached and subsequent volume calculations will be over exaggerated. These factors are important to consider with plethysmography results, particularly in diseased patients as actual contributing gas mixing airways are likely to be smaller than the lung volume would suggest.

1.4 Computed Tomography

As described, the mutation of the CFTR gene and the disease manifestations this causes, increase the likelihood of airway narrowing due to inflammation, or mucus plugging, distal gas trapping, and ultimately bronchiectasis (irreversible structural changes to the conducting airway wall) in the CF lung [67].

High Resolution Computed Tomography (CT) is a scan of the lungs made up of 2D slices that produce very accurate pictures of the structural damage in the lungs, including bronchial wall thickening, mucus plugging and bronchiectasis [68]. Structural abnormalities occur early in CF disease and can be identified by CT [25] even in infants [69]. For CF, chest CT scans are now the gold standard technique for identifying bronchiectasis [70] but CT scores have not only shown the presence of disease [71], they can also monitor changes and have been shown to improve after treatment [25]. Shah *et al.* 1997 completed two CT scans on symptomatic (exacerbating) and asymptomatic CF patients within a two-week period; even in asymptomatic patients there were abnormalities identified on the CT scan. After an exacerbation there were improvements in CT scores but some of the abnormal results did not correspond with improvements in FEV₁ showing that CT can accurately reveal disease that is not detected by physiological outcome measures and offer complimentary information on an individual level [14, 71-73].

Whilst CT scans are very useful, they involve exposure to ionising radiation, which raises concerns for some, particularly in young children. Although CT scores have been shown to correlate well with physiological outcome measures [74] as well as quality of life questionnaires [75] it may be the only outcome measure to exactly identify where specific abnormalities are located [70] so it is likely that the added value of the test will ensure it is used in combination to detect and monitor CF disease. More recently, non-radiation based imaging techniques are being developed, largely based on magnetic resonance imaging (MRI) with or without hyperpolarised gas [76-78]. These techniques are not widely available, are in the early stages of development at our centre, and were not able to be assessed in the work described in this thesis.

1.5 Patient reported outcomes

A patient reported outcome describes health status directly from an individual and assesses how a patient feels about their condition or status

(<https://www.fda.gov/Drugs/DevelopmentApprovalProcess/DrugDevelopmentToolsQualificationProgram/ucm370262.htm#pro>). Patient reported outcomes have clinical efficacy as they directly assess tangible benefit to the patient [79] that cannot be assessed by clinicians or surrogate endpoints. In CF specifically, reported outcomes can give information about respiratory symptoms, triggers for seeking care and the most burdensome attributes of adverse pulmonary exacerbations [80]. An established and recognised output is the respiratory domain in the revised CF questionnaire (CFQR QoL) [81]. It has been shown to improve after a treatment and relate to disease severity [75, 82, 83]. Potential utilities of patient reported outcomes include symptom screening, home monitoring and more recently assess treatment and trial efficacy [84], current utility in clinics is low but there is learning from use of these outcomes from clinical trials.

1.6 Monitoring mild lung disease

Many CF patients now only show subtle changes or in fact have normal lung function when utilising FEV₁ [85]. For example, routine treatments only show an FEV₁ treatment advantage when compared to a controlled placebo group rather than impacting FEV₁ alone [56]. FEV₁ progression, though showing the steepest decline between ages 11 to 15, is fairly stable throughout life [19] and data from the UK CF registry provide information that median FEV₁ has only improved over time [2]. Other factors considered to be useful in the monitoring of CF disease include rate of pulmonary exacerbations and even the variability of the FEV₁ has been shown to give an indication of lung function loss over time [15] but the utility is yet to be shown. It is becoming increasingly difficult to monitor early onset of disease and prevent irreversible damage due to the lack of sensitivity of FEV₁ to detect subtle changes in light of the improved care [45, 86, 87]. Therefore, collaborative efforts to find sensitive outcome measures to quantify and track early disease are paramount [34, 88]. Emerging outcomes that are useful going forward into clinics will have to be explored in a trials setting to independently assess the usefulness of a measure in a controlled research environment. Studies utilising FEV₁ require large study numbers as well as long treatment interventions to be able to detect therapeutic benefits [89]. With the uptake of novel therapeutic treatments, the pool of patients to test who are naïve to new study medications is also decreasing, so study design and providing adequate power of trials may be more complex [34].

1.7 Concept of a minimal clinically important difference

A meaningful improvement in a measure may not simply be a statistically significant change after a treatment but depends upon its magnitude and relationships with baseline characteristics and other outcomes. A minimal clinically important difference (MCID) is the smallest change in a measure that a patient perceives as beneficial [90]. Determining an MCID for an outcome measure is not

straightforward; the change should correlate with change in a recognised outcome measure (an anchor) as well as be larger than the variability of the measure [52, 79]; it is likely that the result is the best estimate of change that would correlate with the experience and intuition of a clinician when using an outcome measure [90]. An MCID for a CF specific patient reported outcome measure has been reported [81] but so far for FEV₁ a definitive MCID is unknown [53].

Factors influencing the identification of an MCID include within test variability, magnitude of treatment responses as well as baseline severity. For FEV₁ short term within-test variability relative to baseline of up to 10% (or 15% in longer observational periods) has been reported [91]. Baseline severity is likely to impact variability and subsequent fall over time, for example short term variability in FEV₁ is higher with more impaired lung function and those with a higher lung function are likely to have larger acute falls in lung function [55]. Table 1.1 shows key interventional trials where FEV₁ has been utilised as the primary outcome (a similar Table has been produced previously by Stanojevic & Ratjen 2016 [89]). The threshold for change in these studies is between 3 and 10% so clearly different treatment interventions, the duration of treatment and the mechanism of action will all impact FEV₁ response. Taking into account the within test variability and change over time, a 15% change in FEV₁ can be confidently described as meaningful [92]. The fact that all of the interventions listed in Table 1.1 went on to be used clinically, despite changes being under 15%, suggests that a 15% change in FEV₁ is too large to be considered the only meaningful threshold of change. As described, other outcomes such as decreasing the pulmonary exacerbation rate will have been considered. That being said, as FEV₁ is more frequently within normal ranges in patients with CF [93], FEV₁ may be seen as a relatively insensitive tool for tracking changes over time.

Interestingly, a recent study looking at the safety and efficacy of combination therapy on 6-11 year olds homozygous for F508del only showed a small improvement in FEV₁ post treatment, despite the treatment being considered safe and showing positive improvements in other measures [94]. This shows that use of FEV₁ in those with mild disease does not provide a useful readout of treatment response and another physiological tool is required.

Table 1.1 Key interventional studies where FEV₁ treatment changes are seen after an intervention

Study	Patient Group	Intervention/ Duration	Absolute FEV ₁ change	p value	Placebo change (natural variability)	Baseline FEV ₁ % (further to change?)	Comments
Retsch-Bogart <i>et al.</i> 2008 [95]	78 CF adults and 27 adolescents	AZLI 75mg or 225mg 14 days	6.9% 6.8%	NS for treatment effect	1.99% diff to placebo in 75mg 0.71% diff in 225mg.No diff	75mg 74.3 (21.8) 225mg 81.23 (17.56)	14 days not significant – maintained and decreased.
Elkins <i>et al.</i> 2006 [30]	164 CF - 82 7% hypertonic saline, 80 0.9% saline	Hypertonic Saline (48 weeks)	3.2%	0.03	No difference	76% with treatment, 73% control.	Treatment group had higher FVC values and fewer pulmonary exacerbations
Fuchs <i>et al.</i> 1994 [29]	968 CF – 157 once daily, 146 twice daily, 325 placebo	DNase (24 weeks)	5.8%	<0.01	No difference	60-61% with treatment, 61% in placebo	Reduced the risk of pulmonary exacerbation
Ramsey <i>et al.</i> 1994 [96]	520 CF -258 Tobramycin, 262 placebo	Tobramycin (20 weeks)	10%	<0.001	2% decline in placebo	49.9% with treatment, 51.2% in placebo	Pseudomonas density decreased with treatment. 26% less hospitalisations. No difference of the effect of tobramycin on disease severity.
Heltshe <i>et al.</i> 2016 [97]	123 CF	IV's for treatment of Pulmonary Exacerbation (least 10 days)	10%	Not reported	n/a	55.8% (20.9)	Best FEV ₁ % in the prior 6months was 63.7 (23.9) 5.9% improvement in children compare to adults (p=0.002) 90% recovered lung function

Study	Patient Group	Intervention/ Duration	Absolute FEV ₁ change	p value	Placebo change (natural variability)	Baseline FEV ₁ % (further to change?)	Comments
Ramsey <i>et al.</i> 2011 [35]	161 (83 ivacaftor, 78 placebo)	Ivacaftor (24 weeks)	10.6%	<0.001	-0.2%	63.5 ivacaftor, 63.7 placebo	Effects noted after 2 weeks and sustained for 48 weeks
Elborn <i>et al.</i> 2016 [98] and Wainwright <i>et al.</i> 2015 [40]	1108 CF (371 placebo,	Ivacaftor and Lumacaftor (400 or 600 mg doses) (24 weeks)	3.3 and 3.7%	0.02 and 0.04	-0.4% in those above 40% FEV ₁	60.4% placebo. 31.1-96.5 across all severity groups in active	Homozygous F508 del Similar absolute changes in FEV ₁ % across severities. Larger relative improvements in those with FEV ₁ % <40%.
De Boeck <i>et al.</i> 2014 [99]	39 CF with non G551D gating mutation >6years (mean 23.8)	Ivacaftor (150mg q12h) 8 weeks	10.7%	<0.0001	-3.2%	78.4 (42.9- 118.7)	Changes detectable after 2 weeks. Open label study FEV ₁ % increased from baseline by 13.5%.

1.8 Multiple breath washout – Lung clearance index

The Multiple breath washout (MBW) technique and its main parameter Lung Clearance Index (LCI) have been proposed as an outcome measure that is more sensitive than spirometry [59] and is able to detect disease abnormality even in those with mild lung disease [100]. It is an easy, non-effort dependent manoeuvre that involves recording a tracer gas whilst it is cleared from the lungs during normal tidal breathing. A key study evaluating CF lung disease progression in 6 – 20 year olds was completed by Kraemer *et al.* 2005 [1]. They were one of the first groups to relate MBW results, specifically LCI, to spirometry measures and assess the deterioration of lung function over time. They found that LCI was the earliest predictor of lung function abnormality and frequently highlighted abnormality in those whose spirometry was within normal ranges. Figure 1.5 shows LCI and FEV₁ Z score data from this paper, the added benefit of utilising LCI as an outcome measure was seen in half of the study subjects (top right hand corner) where LCI was abnormal but this abnormality was not detected by FEV₁. The study showed that MBW testing could provide a more complete assessment of lung function decline and in part improve the understanding of the pathophysiological processes in CF disease. That being said LCI is not yet an established surrogate endpoint for clinical trials. The next section describes the MBW test, its clinimetric properties and steps to establish LCI derived from MBW as a surrogate outcome measure.

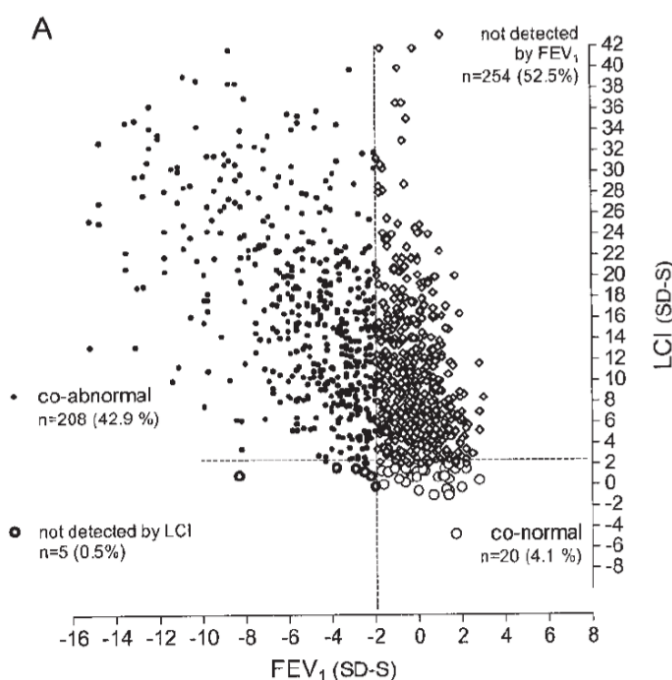


Figure 1.5 Comparison of FEV_1 and LCI Z scores from Kraemer *et al.* 2005 [1]. The dotted lines split the data above and below 2 standard deviations (SD). 4.1% of the group had normal lung function as defined by both FEV_1 and LCI within 2SD (bottom right hand corner). 42.9% had abnormality shown by both FEV_1 and LCI (top left hand corner). Abnormality was not detected by LCI in 0.5% of people (bottom left hand corner) but FEV_1 was unable to detect lung function abnormality picked up by LCI in 52.5% of patients (top right hand corner). Reprinted with permission of the American Thoracic Society. Copyright © 2017 American Thoracic Society. *Annals of the American Thoracic Society* is an official journal of the American Thoracic Society [1].

1.8.1 Ventilation Inhomogeneity

As demonstrated, FEV_1 gives a good indication of flow through the large conducting airways but it lacks sensitivity to test the more distal acinar airways in part due to the nature of the forced expiratory effort; this is a particular issue in those with relatively mild disease. The further down the respiratory tract the more airway bifurcations there are, in turn increasing the surface area of the lung (dotted lines on Figure 1.3). Airflow changes from convection to diffusion and it is here that gas mixing and exchange in the alveoli take place. Gas mixing efficiency can be measured by single breath and multiple breath washout techniques.

In the single breath technique the gas expirogram is broken down into three phases (Figure 1.6). Phase I represents absolute physiological deadspace, Phase II the rapid convection of air and Phase III the slower alveoli gas depletion. Some breath traces may also contain a phase IV stage where there is a rise after expiration due to the closing volume of the airways. An increased phase III slope depicts increased V_I . The single breath washout involves a maximal manoeuvre so it assesses V_I at a different lung volume to the multiple breath washout (MBW) [101]. The physiological deadspace calculation has been described previously [102] and when looking at phase III analysis increased physiological deadspace has been shown to be indicative of the disease process in the acinar region [49]. The MBW technique appears to provide more information about V_I in the smaller airways [101] and may therefore be more useful in the detection of mild disease than the single breath washout. This thesis will concentrate on results from the MBW.

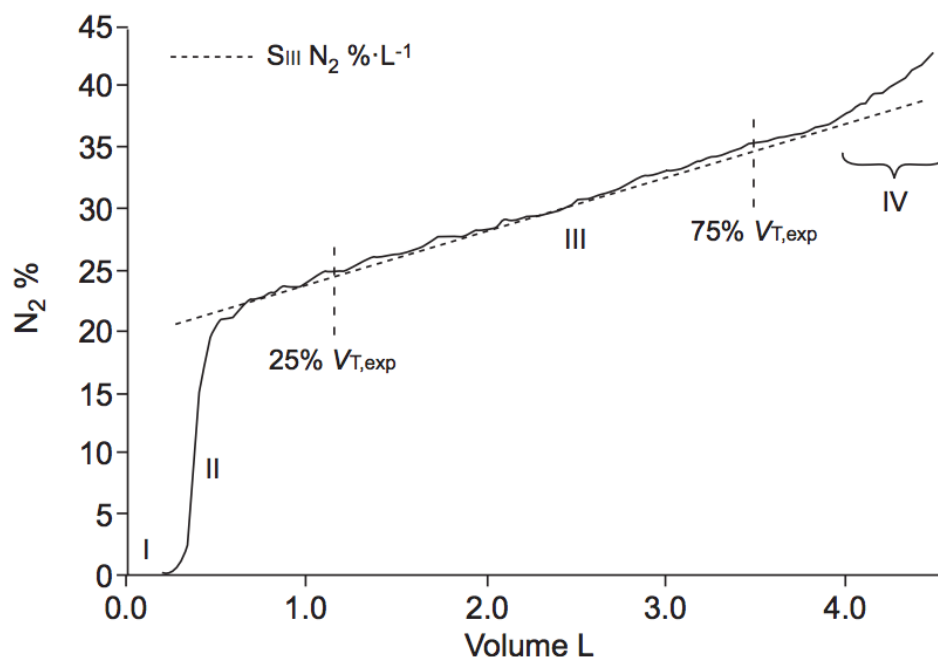


Figure 1.6 A nitrogen expirogram with calculation of the phase III slope (S_{III}) between 25-75% of expired tidal volume. Phase I is the anatomical deadspace between the airways, Phase II the rapid convection of air through the conducting airways and Phase III the diffusion of gases in the alveoli. Reproduced with permission of the European Respiratory Society

©. European Respiratory Journal Mar 2013, 41 (3) 507-522; DOI:

10.1183/09031936.00069712 [101].

1.8.2 Multiple breath washout testing

Multiple breath washout (MBW) was first utilised more than 50 years ago [103, 104] where a quantitative measure of gas mixing efficiency was required for further understanding of lung volume measurement in diseased patients, particularly where gas transport through the lung was thought to be suboptimal. MBW testing enables measurement of lung volumes as well as providing an overall indication of a subject's ventilation heterogeneity and has been described as a sensitive measure of lung physiology [105]. Traditional testing used resident nitrogen (N_2) washed out by 100% oxygen (O_2); it was suggested that the rate at which N_2 was cleared offered insights to the pulmonary structures [106] and in turn the ventilatory process and gas mixing [103].

Renewed popularity of MBW has occurred in the last decade [107] as better gas analysers and recognition of MBW's capacity to determine early lung disease was revealed [59, 100]. The technique was adapted to incorporate truly inert tracer gases i.e. Sulphur hexafluoride (SF_6) or Helium (He) that are washed out by room air. Various studies and further analysis of MBW traces have been completed with SF_6 but the utility of this gas has recently been limited as it has mild anaesthetic properties [108] and SF_6 in large quantities is a potent greenhouse gas [109]; legislation to reduce these man-made gases are now in place in various industrial sectors although the amounts of gas required for MBW are extremely low and regulatory-approved equipment is available.

In an open-circuit system washout gas is newly breathed and blown out of the system whilst the end tidal concentration (Cet) of tracer gas is recorded with each breath. The Cet can be determined in a number of ways but conventionally it is determined by the average of 95-98% of the expired breath volume [101].

MBW can also be performed with a closed circuit system (the gas breathed is contained within a circuit and rebreathed to calculate results i.e. with no requirement of additional gases and a ratio of the contained gases give the results). Most laboratories utilise an open circuit system due to variability in results seen with the closed system perhaps due to hypercapnea with the rebreathing set up [110], although a closed circuit system with a carbon dioxide (CO_2) scrubber has been introduced [111] showing that washout may be quicker and easier to perform in a clinical setting. The rest of this thesis will refer to the MBW test performed with an open-circuit system as it has been most widely used and adopted in clinical and research settings at the RBH.

1.8.2.1 Determining the end of the test

Conventionally, tracer gases are cleared to $1/40^{\text{th}}$ (or 2.5%) of their starting concentration and the test can be terminated once this concentration is achieved after three consecutive breaths [101]. This is an arbitrary value as a factor of the limit of the gas analyzers used traditionally but most groups still use this cut off for standard calculation of LCI. In disease it will take more breaths to clear a tracer gas down to $1/40^{\text{th}}$ and therefore the washout will be longer [47, 112, 113]. This is a factor of poorly ventilated lung units and impaired gas mixing that dominate the washout curve [114] and different regions of lung that ventilate more slowly than others [115].

1.8.2.2 Devices

The gold standard device for MBW has long been regarded as the Respiratory mass spectrometer (RMS). Methodology using 4% SF_6 has been driven by specialist paediatric respiratory centers with expensive gas analyzers and the ability to perform MBW within a research setting [59, 93]. The RMS measures gases directly but it is large, non-portable and complex to maintain. A number of alternative gas analyser devices are now available and new technology has enabled the use of N_2 as a tracer gas as well as 0.2% SF_6 with more widespread adoption of the technique [116, 117]. These devices include the modified Innocor device (Innovision) using SF_6 and the Exhalyzer D (Ecomedics) and EasyOne Pro (NDD technologies) using N_2 . Detailed explanations of each device are provided in the methods section.

1.8.3 Parameters derived from the Multiple Breath Washout test

1.8.3.1 Lung Clearance Index

The main parameter calculated from MBW is LCI. LCI is a robust ratio of the cumulative expired volume (CEV) (total expired gas to clear the tracer) divided by functional residual capacity (FRC) ($\text{LCI} = \text{CEV} / \text{FRC}$).

By dividing by FRC the result is corrected for lung size, providing a widely applicable unitless number. FRC is the volume of gas remaining in the lung after a normal, tidal expiration. It is calculated by dividing the net expired volume divided by the Cet at the start, minus the Cet at the end ($\text{FRC} = \text{net volume of expired gas over the washout} / (\text{Cet start} - \text{Cet end})$). As MBW is limited to only communicating lung units, FRC values may be lower than FRC from plethysmography, particularly in disease where bronchiectasis, airway narrowing and mucus

plugs cause obstruction so areas of lung may not be reached.

CEV is the total volume of air expired over the washout, calculated using a flowmeter on each device.

Thus, increasing ventilation heterogeneity will result in a higher LCI. This is because the gas mixing within the lungs is more variable with disease and therefore requires more volume to get down to the target C_{et} revealing the degree of disease severity in a higher value. Healthy control (HC) LCI values have been shown to be between 5 and 7.5 [59, 93, 105, 118, 119], but in CF more lung turnovers are required to clear the tracer gas because of airway mucus and narrowing and remodelling of the airway walls (ultimately, bronchiectasis) related to inflammation and infection. Values in CF can reach up to and above 15 [105]. Table 4.1 shows different studies, utilising different equipment but generating HC and CF patient LCI results. There seems to be little influence of age, except below 6 years where there is an inverse relationship to height; LCI is higher at shorter height and in turn smaller lung size [120]. This will not be looked into further since in this thesis only patients above 6 years old have been explored; but the results are important to note and work including but not limited to Great Ormond Street hospital/Institute of Child health and the Australian Respiratory Early Surveillance Team should be referred to for more comprehensive information. There may also be an upper limit of MBW parameters when further analysis of the MBW test in more severe disease is completed [121] but as this thesis will not utilise further parameters of the MBW test this should not be of concern.

1.8.3.1.1 What is LCI measuring?

Ventilation distribution is essentially split into two compartments: convection dependent inhomogeneity (CDI) (the non gas exchanging lung zone in the conducting airways) and the diffusion convection dependent inhomogeneity (DCDI) (the gas exchanging lung zone in the periphery/acinar region) [49, 50, 101]. LCI is more affected by CDI, as it is a larger proportion of the washout process. CDI or specific ventilation are affected by lung compliance and resistance of the supplying airway [112], essentially computing asynchrony of convective flow for each lung unit which in turn makes LCI lower (good convection) or higher (more flow heterogeneity between lung units). DCDI also contributes to the LCI signal as it is a whole lung measure and is thought to relate to the gas diffusion asymmetries or structural changes around or beyond the diffusion convection front (acinar airways) [49].

LCI will increase with obstruction from secretions and airway inflammation, which is thought to change transiently over days and this, may be shown by physiotherapy studies [122] or

treatment of acute exacerbations [123, 124]. High LCI will also be influenced by structural changes as shown by HRCT scores [125]. Only with longitudinal or targeted studies to see transient changes of LCI will the variability of LCI be revealed. As LCI is a global measure of lung function it is difficult to tell if these parameters are distinguishable when using LCI alone. It is likely that other parameters from the MBW trace or complementary physiological outcome measures would have to be utilised.

1.8.3.2 Further analysis of the MBW trace

As discussed, LCI is a marker of whole lung ventilation and the two different compartments CDI and DCDI. If LCI was within normal ranges or the location of the VI was important i.e. acinar or conducting airways then the trace can be explored further by four methods (Phase III slope analysis [126], Moment Ratios [127], Curvilinearity [128] and Slow and Fast Moving lung units [129]). Essentially all four methods split the washout into sections and the weight of the predominant contributing lung unit gives us an indication of where the most VI is occurring. Whilst these parameters are utilised [127, 130, 131] and have proved to distinguish between CF and non CF [121], LCI is the most reported in the literature and some suggest that MR do not appear to add any additional information [47].

The most widely used additional MBW parameter is Phase III analysis. The sum of the phase III slope from multiple breaths shows the magnitude and progression of VI, it is increased with increased inhomogeneity. By splitting the slope into two the acinar (Sacin which reveals DCDI) and conductive (Scond which reveals CDI) ventilation can be shown [114, 121, 128]. A study found larger Sacin results in CF compared to both asthma and healthy controls, attributing more VI to the gas exchange zone in these subjects [132]. But Sacin and Scond are yet to show much clinical utility, likely to do with the complexities of calculation as well as differences in more advanced disease [121], so the benefit of understanding the region where VI is most prominent is unknown [133].

Curvilinearity (curv) and Moment ratios both seek to explain the difference between the quick decay of the tracer gas seen initially in the washout trace vs. the release of gas by poorly ventilated areas i.e. the long tail seen in more severe patients. Conceptually curv describes the first half of the washout compared to the latter half, if they are equal then the value would be zero, whereas in disease the value is closer to 1. Like LCI, Curvilinearity looks at the whole washout curve but unlike LCI has been suggested to provide a large signal to estimate the proportion of lung that is creating the most heterogeneity [126].

Moment ratios are more complex but essentially a weighted plot of the proportion of tracer released in the latter stages is calculated from the end tidal concentration vs. turnover [133]. Curv was able to distinguish between health and disease, it correlated well with LCI and attributed abnormality of a washout to CDI (40-50%) and DCDI [128]. Moment ratios are inline with LCI results [134] but are largely impacted by breathing pattern so results may be variable [130].

The MBW test has many different parameters that could potentially reveal far more about the pattern of the washout trace and heterogeneity of the expired tracer gas than LCI alone [128], yet there is growing evidence that some of the mathematical theory may be flawed [121]. Verbanck *et al.* 2013 have updated the mathematics of the equation i.e. using a different cut off (Sacin* and Scond*) and suggest that the plateau seen in more severe patients has now been rectified [114]. But Horsley & Wild (2015) suggest it is too early to predict how useful these measures will prove to be [107].

Clinical utility of further MBW analysis may be limited by the complexity of analysis and the lack of studies correlating these results to known outcome measures [133]. Phase III has also not shown a consistent response to treatment [135, 136]. As this thesis concentrates on the clinimetric properties of LCI and taking steps to introduce LCI as a surrogate endpoint in clinical trials, the further parameters available from MBW have not be assessed.

1.9 Clinimetric properties of Lung Clearance Index

To gain acceptance as an endpoint for clinical trials, LCI must fulfil stringent criteria [52]. Outcomes must reflect the presence and severity of disease, be reproducible and show improvement after a treatment that match changes seen by the true endpoint in question [25]. Kent *et al.* 2013 systematically reviewed the LCI literature to assess and show evidence of LCI's clinimetric properties [52]. They concluded that LCI was an attractive endpoint for multicentre use, particularly for children where FEV₁ is in the normal range, and had useful clinimetric properties including being feasible, reliable and valid.

From the review, there were two outstanding themes to establish LCI as a surrogate endpoint for clinical trials: 1. That further comparisons of different MBW methods should be completed and 2. That further information about LCI's responsiveness to treatment and correlation to other clinical outcome parameters should be collated.

This review section will summarise the main clinimetric properties of LCI and build on the work by Kent *et al.* 2013.

1.9.1 Feasibility

As a tidal breathing test MBW has been shown to be feasible in non-sedated children as young as 2-3 years [100]. With the incorporation of distraction techniques and familiarisation to the test, 79% of preschool children were able to complete 3 technically acceptable MBW traces [100]. The importance of operator knowledge and recognition of good quality online/real time analysis is also important for LCI feasibility [133]. Two studies have tested young children with respiratory disease to assess the feasibility of MBW testing during clinic settings. The first study used a set 20-minute time slot to try to achieve the targeted triplicate MBW tests all within acceptability criteria outlined in the consensus statement. Only 24% of children were able to achieve 3 adequate tests, but 60% did achieve two tests [137]. This has led to subsequent studies looking at the feasibility of two rather than three MBW trials within a testing session [138]. Providing there is sufficient experience within a testing site, LCI results from two tests do not seem to change the mean result or compromise sensitivity, but triplicate results are still the standard. In another study, 70% of children were able to achieve acceptable MBW tests confirming that the majority of novice children are able to perform this technique [139]. These studies suggest that there is an increased success rate with increased age and therefore specific preschool testing guidelines are warranted. The fact that both studies were able to complete the testing during a routine clinic is promising. Research visits are generally less limited in time and have dedicated research staff so the visits and tests can be completed with ease. If MBW is achievable in clinic then this may lead to increased use [140, 141].

Adult MBW testing is considered feasible from achievability and cooperation aspects and many studies have had high success rates in adults [105, 142, 143]. With age there may be increased disease severity, which increases the duration required on the machine [101, 144], requiring up to 1 hour; therefore the practicality and feasibility of completing MBW in these patients is questionable. Whether the additional information from MBW is beneficial in these patients should be explored further.

1.9.2 Validity

Validity can be described as the degree to which a test correlates with another test that measures the same attribute; it is important to know how LCI can discriminate between groups and compare to the gold standard of lung function testing [52]. LCI has been shown to be able to discriminate between individuals with CF and healthy controls [118, 145-147]. Aurora *et al.* 2005 found that LCI was able to detect abnormality in more preschool patients

than plethysmography and spirometry from those who were able to complete it [100] and all of the measures were higher in CF compared to HC. Stahl *et al.* (2014) took this further and performed MBW measurements on 14 CF infants aged 0-4 years and 26 aged matched controls at routine clinical annual assessments [145]. The CF infants had a significantly higher LCI than the healthy control infants, showing the ability of the test to detect changes in ventilation heterogeneity even at this early age. This is significant in that spirometry cannot be performed with this population so LCI is an appropriate alternative to determine the state of the airways. The fact that the tests were incorporated into a clinical annual assessment is important, it supports the diagnostic monitoring of early lung disease within clinical management.

LCI has also been able to detect more subtle changes in disease i.e. age [86] where groups have looked at LCI in young children [120] and over a lifetime [143] and although the global measure of VI is relatively stable it is not independent of age. LCI is even able to detect differences in infection status and structural changes in CT scores [148] and been shown to detect changes in CF patients with different phenotypes i.e. children vs. adults [105] or patients with or without *Pseudomonas aeruginosa* infection [93].

1.9.3 Reliability

Reliability is the ability of a system to perform a measurement within required limits, consistently and free from error [52]; tests that are reliable are accurate, repeatable and consistent within and between tests. Kent *et al.* 2013 report that the majority of MBW studies within their meta-analysis had low within test coefficient of variation (CV) between 3 and 7% [138, 141, 149]. One study even found that between test and between site variability was low, suggesting that LCI results are reliable and applicable to a wider population [149], this has also been found in other studies [150-153] and even in preschool patients [100]. Determining global healthy control reference ranges will undoubtedly add to the robustness of the LCI test in turn increasing the reliability of testing within and between centres.

1.9.4 LCI and relationship with other outcomes

To start to understand what an LCI result means in relation to disease status and clinical stability we need to compare it to other available physiological biomarkers. Multiple studies have confirmed LCI is more sensitive than spirometry in early disease [45, 74] also proving to be the earliest detectable and most predictive measure of disease progression when annual

follow up of 142 children was performed [1]. Advances in therapy have led to many people with CF having FEV₁ within normal ranges [45, 154]. LCI is a reproducible tool for detecting subtle changes in the small airways [93, 100] and this would indicate that MBW could be used as a surrogate endpoint for these individuals.

High-resolution CT offers direct visualisation of structural changes relevant for CF disease progression [25] but carries inherent radiation risks. LCI correlated better with CT than with spirometry [74, 148], reflecting the extent of bronchiectasis and air trapping within CT scores [71]. This would suggest that MBW is suitable non-invasive tool for characteristic structural changes within the lung in CF disease.

Vermeulen *et al.* (2013) highlight that to validate a physiological measurement as a surrogate endpoint it should be correlated with validated clinical outcomes such as pulmonary exacerbations or patient reported outcomes [87]. 63 CF patients were prospectively followed over a year after baseline and subsequent FEV₁, LCI, CFQR QoL scores [84] and pulmonary exacerbations were recorded. In patients with normal spirometry LCI was more closely related to pulmonary exacerbations and CFQR QoL than any other variable. They concluded that this highlights the superiority of LCI over FEV₁ and therefore in the early stages of CF lung disease LCI should be utilised more readily. They suggest that the use of LCI as an outcome requires more refinement i.e. knowledge of the effect of exacerbations on LCI and the timing and magnitude of LCI improvement with interventions to determine what represents a clinically meaningful change.

1.9.5 LCI treatment responses

Studies have looked at LCI changes before and after treatment of an acute exacerbation [124, 151] and have found that LCI improves after IV antibiotics. Whether the improvement was 'clinically meaningful' was not assessed [124]. Singer *et al.* 2013 report that a physiologically meaningful change is likely to be above 1 LCI unit, as this was above the variability they found in paediatric CF patients [141]. However not all studies yield a positive LCI response and there is variability between patients [136, 155-157]. Explanation of this may simply be responders and non-responders to treatment. It may also reflect the heterogeneous and paradoxical nature of LCI where previously non-contributing airways now add to the gas mixing but they are still poorly ventilating which in turn extends the washout [151]. This can be shown in physiotherapy studies where LCI results are variable [122, 158] and yet we know that physiotherapy is one of the mainstays of CF treatment with clear benefits in FEV₁ [122]. Perhaps the benefits of LCI with regard to pulmonary exacerbations only exist in specific patients i.e. milder disease or responders to treatment or

perhaps rather than looking acutely, MBW should be viewed over time.

LCI is beginning to be used in clinical trials of both established [152, 159] and novel therapies [38]. There are substantial differences between the treatments although LCI is able to detect the response to several treatments (Table 1.2).

LCI significantly improved after 4 weeks of hypertonic saline in paediatric patients with CF and normal spirometry [159] as well as a similar study in dornase alpha [152]. Despite overall results being significant, individual results were variable and did not change in a positive direction for some [153] again showing that LCI may be susceptible to mucus shifts and or the opening of previous poorly ventilated/blocked lung units which can dilute the picture.

Initial ivacaftor studies showed reduced sweat chloride response as well as improved FEV₁ in patients who were in the abnormal range at baseline [35]. LCI was used as the primary outcome in a recent study of clinical response to ivacaftor in CF patients with G551D-CFTR mutation and spirometry above 90% predicted [38]. Compared to the placebo group a mean decrease in LCI of -2.16 ($p < 0.0001$) was found, showing decreased ventilation heterogeneity after 28 days of ivacaftor. Despite only mild lung disease manifestations, LCI was responsive enough to detect a change in treatment after just a shorter time period, identifying a significant advantage over FEV₁.

LCI can show benefits over currently available outcome measures but the responses are heterogeneous. Loeve *et al.* 2013 suggests that further information about LCI's ability to show improvements and reflect that of the true endpoint are required [25]. Defining the patient group where LCI is most applicable and what a meaningful clinically important change is are required before LCI can be established as a definitive surrogate endpoint.

Table 1.2 LCI treatment changes after an intervention (IV antibiotics).

Study	Patient Group	Intervention/ Duration	LCI change (SD)	p value	Placebo change (natural variability)	Baseline LCI (SD or range)	MBW device and gas	Comments
Horsley <i>et al.</i> 2013 [151]	44 CF adults	IV antibiotics (clinician decision)	-0.8 (1.4) units	0.003	n/a	14.6 (2.7)	Modified Innocor 0.2% SF ₆	LCI fell in 22 (69%) of subjects. No patients below the ULN for LCI. FEV ₁ also improved significantly (0.32 (0.48), p=0.0006).
Vanderhelst <i>et al.</i> 2014 [136]	15 CF adults	IV antibiotics (clinician decision)	-0.5 (4.5%)	0.045	n/a	10.6 (1.5) to 10.1 (1.6)	Morgan analyser N ₂	A decrease did not occur in all patients. The decrease was associated with Sacin decrease. 1 L tidal breathing. FEV ₁ also improved significantly.
Welsh <i>et al.</i> 2014 [156]	27 CF children and adolescents	IV antibiotics (clinician decision)	-0.53 (4.4%)	NS 0.14	n/a	12.18 to 11.65	Exhalyzer D N ₂	LCI improved in 15 (55% of patients) compared to FEV ₁ where 18 (67%) improved (-3.05 to -2.86 z scores) but was not significant p=0.84.
Yammine <i>et al.</i> 2014 [157]	19 CF children	IV antibiotics (clinician decision)	-0.3	NS 0.41	n/a	13.2 to 12.9	Exhalyzer D N ₂	7 out of 23 clear improvement (>1 unit) FEV ₁ improved significantly.
Robinson <i>et al.</i> 2009 [155]	28 CF children and adolescents	IV antibiotics (clinician decision)	-0.48 (-3.8%)	0.03	n/a	10.10 (6.87- 14.83) to 9.62 (7.37-13.45)	Vmax 22D and Sectra software N ₂	Heterogeneity of the response – improved in 21 of 28 patients. FEV ₁ increased by 7% (0.11L) p<0.01 it improved in 21 of 28 subjects.

LCI treatment changes after an intervention (Hypertonic saline and DNase)

Study	Patient Group	Intervention/ Duration	LCI change	p value	Placebo change (natural variability)	Baseline LCI (SD or range)	MBW device and gas	Comments
Amin <i>et al.</i> 2010 [159]	20 CF children and adolescents	Hypertonic Saline (4 weeks)	-1.16 (0.26 to 2.05)	0.016	Cross over design	7.86 (1.71) to 8.89 (2.10)	Mass Spectrometer 4% SF ₆	There was no difference between either group at baseline. The majority of patients decreased their LCI. FEV ₁ (62ml improvement but variability high) and CFQR were not significant.
Amin <i>et al.</i> 2016 [160]	16 CF children and adults	Hypertonic saline (single dose 24 hours)	-0.6 (0.32)	NS 0.08	Cross over design	12.2 (6.5-21.1)	Exhalyzer D N ₂ Mass Spectrometer 4% SF ₆	SF ₆ LCI -0.59 (0.5) p=0.24. Trend for decrease, more variability in SF ₆ RMS. Repeated measures analysis (5 time points in 12hours = ns). FEV ₁ improved by 0.34(1.96) p=ns – baseline 86.9 (47.4-102.8).
Subbarao <i>et al.</i> 2013 [153]	25 CF children (under 6 years)	Hypertonic saline (n=12) (48 weeks)	1.43 (-0.13 to 2.99)	0.073	Isotonic saline (n=13)	UNK	Mass Spectrometer 4% SF ₆	Infant study of inhaled Saline. Treatment effect was significant when using Z scores (account for height). Those with abnormal LCI improved. FEV _{0.5} showed significant improvements [161]
Amin <i>et al.</i> 2011 [152]	17 children and adolescents	DNase (4 weeks)	-0.9 (1.44)	0.022	Cross over design	6.13 (0.41)	Mass Spectrometer 4% SF ₆	All patients had abnormal LCI. Within test CV% was 6.8 ULN 13.6% - it did not relate to the magnitude of LCI. The placebo group also decreased LCI. MEF ₂₅₋₇₅ also improved by 6.1% (0.28 z scores) but FEV ₁ did not.

LCI treatment changes after an intervention (Ivacaftor)

Study	Patient Group	Intervention/ Duration	LCI change	p value	Placebo change (natural variability)	Baseline LCI (SD or range)	MBW device and gas	Comments
Davies <i>et al.</i> 2013 [38]	17 CF with FEV ₁ % >90	Ivacaftor 28 days	-2.16 (-2.88 to -1.44)	<0.0001	Cross over design	8.88 - 9.17 at the start of treatment for both groups.	Modified Innocor 0.2% SF ₆	Baseline LCI 7.4 or above and FEV ₁ above 90% – G551D on at least one allele. All patients improved LCI and had responded by day 15. FEV ₁ improved (8.67% (2.36 to 14.97, p=0.01).
Lenherr <i>et al.</i> 2015 [162]	1 CF child	Ivacaftor (41 weeks)	-6.7	n/a	n/a	14.5 to 7.8 after 41 weeks	Exhalyzer D N ₂	Genotype – S549R/1717-1G. FEV ₁ remained in the normal range (-1.2 to 0.5 z scores).
Harman <i>et al.</i> 2014 [135] (abstract)	8 CF children and adults	Ivacaftor (9-12 months)	Improved	UNK	n/a	UNK	Modified Innocor 0.2% SF ₆	G551D mutation. Improved despite starting LCI. Phase III slope did not add any additional information. FEV ₁ improved from by 11.4%
Ratjen <i>et al.</i> 2017 [94]	102 placebo 104 Ivacaftor and Lumacaftor	Lumacaftor and Ivacaftor 24 weeks	-1.0 (-1.3 to -0.8)	p<0.0001	0.1 (-0.2 to 0.3) p=0.5390	10.3 (2.3)	Exhalyzer D N ₂	F508del patients. LCI had improved by day 15. Response larger than CV%. FEV ₁ also improved but to a smaller degree (2.4ml (0.4 to 4.4) p=0.0182).

LCI treatment changes after an intervention (Physiotherapy)

Study	Patient Group	Intervention/ Duration	LCI change	p value	Placebo change (natural variability)	Baseline LCI (SD or range)	MBW device and gas	Comments
Fuchs <i>et al.</i> 2010 [163]	27 CF children and adolescents	Physiotherapy (1 hour)	-0.2 (-0.51 to 0.11)	ns	Yes - no intervention	7.76 (1.23) to 7.96 (1.04)	EasyOne Pro NDD 4% SF ₆	Lower LCI in intervention group= NS. No short term effects. LCI not consistent. Within test CV% good but higher between tests in CF compared to HC. -0.05 (-0.47 to 0.37 FEV ₁ Z Score = NS). Physiotherapy – endurance training, thorax mobilisation and airway clearance (flutter or PEP).
Pfleger <i>et al.</i> 2015 [122]	29 CF children and adults	Physiotherapy (30 minutes)	-0.2 (-4.84 to 3.37)	ns 0.6	n/a	15 (7.4-24.2) to 14.8 (7.14- 22.39)	Exhalyzer D N ₂	LCI improved in 18 patients and deteriorated in 11 people. LCI is unpredictable. LCI CV% was low (2.95 and 3.2%). FEV ₁ improved 70.04 (29.8-115.8) to 72.28 (33.9-119) = 2.24 (-8 to 16.1) p<0.01. PEP mask physiotherapy
Grosse- Onnebrink <i>et al.</i> [164]	20 CF children and adults	Physiotherapy	-0.9 (-0.45 to 3.47)	0.002	21 CF control group no intervention	11.8 (8.81 to 19.28) to 11.1 (8.15 to 17.8).	Spiroson NDD 4% SF ₆	LCI decreased in 15 out of 20. High coefficient of repeatability. High frequency chest wall oscillation physiotherapy. No differences between groups.

1.10 Standardisation of MBW

MBW working groups have formed to try to standardise MBW as an outcome measure not only for clinical trials, but also clinical care in CF [52, 101, 116]. All age assessment of LCI should be completed and global regulatory agency approvals need to be sought [116]. Whilst the majority of MBW devices have undergone varying degrees of validation and patient testing and equipment requirements are understood [101], data arising confirm that results are not interchangeable [118, 165]; evidence supporting this has been further gathered as part of this thesis. Other practical elements should also be addressed such as patient interface, test duration as influenced by trial number [166], cut off ranges [167], transparent raw data [133] as well as further exploration of differences between test gas selection [118]. If LCI is to become a useful outcome measure for multicentre clinical trials, standardisation across sites and collaborative multicentre efforts are necessary.

1.10.1 Guidelines

In 2013, Robinson *et al.* produced a consensus statement endorsed by the ERS and ATS on inert gas washout measurement [101]. The manuscript provides information about equipment specifications, test performance and analysis in light of renewed interest in this technique. The expert commentary provides clear review and instruction on multiple breath washout testing and takes steps to highlight important issues and questions that remain unanswered. The authors suggest that manufacturers should work with researchers to ensure equipment is robust and validated which will lead to integration of the test into routine clinical care.

Following on from this, a detailed standard operating procedure was published for the N₂ Exhalyzer D (http://www.mbwtraining.com/ECFS_MBW_SOP.pdf, Jensen *et al.* 2013).

The European Cystic Fibrosis Society, Clinical Trials Network (ECFS CTN) brings together experienced CF centres throughout Europe with an aim to intensify CF clinical research and bring new medications to patients as quickly as possible (<https://www.ecfs.eu/ctn>). One avenue of this aim is to improve outcome measures for both clinical and research use by a process including rigorous standardisation.

The group have adopted the standard operating procedure by Jensen *et al.* (2013). In 2014 an LCI core facility was established at the Royal Brompton Hospital (Jane Davies, Clare Saunders, Christopher Short and Katie Bayfield) with aims to certify sites and operators after standardised training, provide centralised over reading to improve the consistency of results and to facilitate global trials with use of LCI as an outcome measure [168]. These efforts

have been made in connection with other core facilities (Hospital for Sick Children, Toronto, Canada (Felix Ratjen, Sanja Stanojevic, Renee Jensen *et al.*) & Westmead Hospital, Sydney, Australia (Paul Robinson *et al.*)) emphasising collaborative efforts to improve LCI as an outcome measure.

Despite these efforts there are still gaps in our understanding, particularly what is a clinically meaningful change in LCI and the impact of utilising different test gases on the results.

1.10.2 SF₆ and N₂

MBW using SF₆ as a tracer gas has received the most attention with a large range of studies describing the benefits of LCI over traditional lung function measures [169], correlation to structural abnormalities [74] and response to treatment [38, 159]. Many factors including expense of gold standard equipment and test gases [170], stricter guidelines on the use of large quantities of SF₆ and availability of tracer gas at each site mean that MBW using N₂ may be utilised more readily going forward. Improvement in N₂ sensors, and new commercially available devices utilising N₂ [116] with useful user interfaces without the need for separate offline analysis (ExD/N₂ and NDD EasyOne Pro systems) also play a role.

1.10.2.1 SF₆

Sulphur Hexafluoride (SF₆) is a heavy and dense gas with a low viscosity as its atomic structure is very stable. Outside forces or changing ambient conditions do not have large impacts on SF₆; this and the fact that SF₆ is an inert gas (not produced or taken up by the body), make SF₆ a very useful tracer in MBW testing. SF₆ is a manmade gas mainly utilised in the power industry for its non toxic, non flammable and insulation properties [109]. SF₆ has a long half-life of approximately 3200 years and concentrations of atmospheric SF₆ have doubled since the 1970's with links to climate change [171]. However the small concentrations utilised in the MBW technique are safe to use and should not be limited for use.

1.10.2.2 N₂

Nitrogen is approximately 78% of air; it is denser than O₂ but not SF₆. N₂ does not take part in gas exchange but there is dissolved N₂ [172] excreted from body tissues [172, 173]. It is a fairly light gas with a molar mass of 28 g/mol and therefore it has been suggested that its diffusion convection front is located more proximally than heavier gases such as SF₆ [47] in that diffusion rate of a gas is inversely proportional to the square root of the molar mass.

Because N_2 is endogenous it is ideal to wash out of the body for this type of testing but there have been several articles to suggest the tissue contributions of N_2 mean that corrections need to be taken into consideration [173, 174]. N_2 excretion from the body can be estimated from tables of exponential equations but the result after correcting for N_2 does not vary by much [175]. A simple equation was described when looking at 7 minutes of 100% O_2 exposure and relating the changes at the end of the washout to body surface area [176]. A recent study by Yammine *et al.* 2015 explained that because of the different time constants and different elimination rates of N_2 contributing compartments, a simple N_2 correction was not necessarily possible [177]. It is likely that N_2 is released from the body in three phases (rapid, 1.5 minutes, medium, 12-15 minutes and slow, 100-200 minutes) corresponding to highly vascular tissues, muscles then fat [173], but the total amount of N_2 is relatively unknown [101]. The simple correction by Yammine *et al.* (2015) was critiqued for being a mainly mathematical application of a physiological concept [178]. It was also stated that N_2 was not likely to be correctable throughout the whole washout as the N_2 release is difficult to measure and may have a higher relative contribution to the washout towards the end [174]. Newth *et al.* (1997) suggest that the N_2 release is not negligible however and should be corrected for [175]. Current MBW guidelines indicate that quantifying N_2 contribution is difficult due to limited knowledge of the baseline N_2 release [173]. In this thesis, an N_2 correction factor will not be applied. The availability of 100% O_2 within hospital settings and the resident nature of N_2 make N_2 a viable option for more widespread use of N_2 MBW and the increased availability of commercial devices reflects this.

1.10.2.3 SF_6 and N_2 MBW comparison

MBW results generated from SF_6 or N_2 do not yield similar results. Jensen *et al.* (2013) completed the first comparison of the two test gases used in MBW in light of the newly commercially available Exhalyzer D device being utilised more readily [118]. They found that N_2 results were higher than SF_6 and began to offer potential mechanisms for the discrepancy.

The speculative mechanisms behind the differences between N_2 and SF_6 LCI results fall into 4 categories.

- Individual differences between washout systems i.e. breathing pattern,
- Equipment differences,
- Tissue N_2 contributing to a N_2 washout signal,
- Inherent gas differences; tracer and washout gas.

It is important to understand differences between devices, particularly as there is no standardised MBW technique and as it stands N₂ and SF₆ are both likely to be utilised going forward. If the results are not interchangeable then results should not be reviewed as such so that misinterpretation of data does not occur. Only with other comparison studies are the N₂ and SF₆ differences likely to be further explained and actual mechanisms behind the discrepancies revealed.

This introduction chapter has highlighted the potential advantages of LCI use over spirometry; it has also highlighted the known clinimetric properties of LCI and yet there are still unknowns and further evidence is required to promote LCI as a surrogate endpoint for clinical trials. This thesis will concentrate on the differences seen between devices; specifically further defining the mechanisms behind differences between test gases, as well as LCI responses to different treatment interventions. Specific hypotheses, aims and objectives are highlighted at the beginning of each Chapter.

2 Chapter 2 - Methodology

2.1 Study Subjects

Participants aged 6 and upward were recruited for work in this thesis. Subjects with CF had a confirmed diagnosis; two confirmed disease-causing genetic mutations and/ or a positive sweat chloride result alongside a compatible clinical picture or positive newborn screening test [23]. HC subjects, with no previous history of airways disease or smoking were recruited from patients' family members or staff allied to Imperial College London or the Royal Brompton Hospital.

Written Informed consent was obtained from each participant; parents/guardians provided consent for healthy control (HC) or CF children aged 6 and upward, with age-appropriate assent from the child.

Ethical approval was obtained from the national research ethics service NRES (Brighton and Sussex Committee) under the umbrella title of "Non-invasive Physiological Biomarkers" study. The aim of this study was to gain longitudinal information on healthy control and respiratory disease patients to identify sensitive outcome measures for disease monitoring. Separate ethical approval for the CF gene therapy multi-dose trial (MDT) was gained from the gene therapy advisory committee (GTAC) and medicines and healthcare product regulatory agency (MHRA).

CF patients were largely recruited from the Royal Brompton Hospital; patients were approached at CF clinics or inpatient wards (both adult and paediatric) but also approached prior to clinic visits via phone or email after seeing upcoming appointments on the hospital electronic database. MDT patients were also recruited from two hospitals in Edinburgh and referred from patient identification centres across the UK. Further details for this trial specifically can be found in Chapter 5. Study visits took place within and were supported by the National Institute for Health Research Respiratory Disease Biomedical Research Unit at the Royal Brompton and Harefield NHS Foundation Trust and Imperial College London. For the treatment section (Chapter 6) of this thesis patients were approached as their clinical team had decided they should start a new treatment. All recruitment in this section was on an *ad hoc* basis and only when a patient was already attending the Royal Brompton Hospital for clinical purposes. For this reason the repeat visits were not scheduled at regular intervals, but instead arranged for the convenience of the patients. This is also the reason for shorter research visits where the ideal three MBW tests were not always performed, as patients were not able to stay for longer. The final two years of my PhD after the MDT were

sponsored by a non-restricted grant from Novartis and Research grants from Professor Jane Davies.

2.2 Demographic Data

Subject demographics were collected at each visit including age, gender, ethnicity, height and weight.

Height was collected in centimetres to one decimal place, following the MDT standard operating procedure (SOP) using a calibrated stadiometer. The subject stood, with their shoes removed, against the stadiometer, their eyes and ears aligned to ensure measurement from the highest point on the head. The subject took a deep breath and their height was taken.

Weight was collected in kilograms to one decimal place, following the MDT SOP using calibrated weighing scales. The subject stood stationary, with shoes, coats and items from pockets removed, on the weighing scales until the scales balanced and a result was given.

2.3 Outcome measures

2.3.1 Spirometry

Spirometry is a mainstay of lung disease classification and the interpretation of spirometry manoeuvres is crucial for clinical correlation with disease progression [58].

Three different spirometers were used to obtain spirometry measures during this project.

The EasyOne World spirometer (NDD Medizintechnik AG, Zurich; <http://www.ndd.ch>) was used in the MDT, whereas spirometry data obtained from clinic or annual review days was collected using the Vitalograph Alpha Touch (Vitalograph, Buckinghamshire; <https://vitalograph.com/products/spirometers/desktop>) or the Masterscreen, San Diego; <http://www.carefusion.co.uk/medical-products/respiratory/cardio-pulmonary-diagnostics/pulmonary-function-testing/body-plethysmography/masterscreen-body.aspx>).

With spirometry there is evidence to suggest that equipment used does not affect the final clinical result [179], particularly as a properly performed spirometry manoeuvre is effort dependent and a simple measurement of flow that can be accurately read by both handheld and traditional lung function equipment. Therefore the use of different devices in this thesis was considered appropriate.

Calibration of the device was performed prior to each spirometry test, i.e. testing the flow rates with a known volume e.g. a 3L syringe, or ensuring the ultrasonic flow device was not damaged. If the flow rates were not within 3% of the expected result then calibration was

repeated [54]. Biological controls of the spirometry equipment were also performed to ensure that deviation or drift of the device was not apparent [54].

Spirometry was performed in accordance with the ERS and ATS guidelines, as summarised by Miller *et al.* 2005 [54] (Table 2.1 and 2.2). A maximal forceful manoeuvre was performed and the best value for forced expiratory volume in one second (FEV₁) or forced vital capacity (FVC) from three tests, within 150ml of each other, was taken as the final result. The spirometry manoeuvre also produces volumes that are said to relate to the small airways (MEF₂₅, MEF₅₀, MEF₂₅₋₇₅, MEF₇₅). Each spirometer reports these values alongside the FEV₁ result.

The GLI equations are global multi-ethnic reference equations that span all ages [57, 63, 180]. To generate the results, in brief, a macro is downloaded (<http://www.ers-education.org/guidelines/global-lung-function-initiative/tools.aspx>) into Microsoft Excel and a reference database is then linked to the subject characteristics once they are input into the Add-In calculator and %predicted and Z score data are generated.

Table 2.1 European Respiratory Society /American Thoracic Society, a summary of spirometry guidelines [54].

Check the spirometer calibration
Explain the test
Prepare the subject
Ask about smoking, recent illness, medication use, <i>etc.</i>
Measure weight and height without shoes
Wash hands
Instruct and demonstrate the test to the subject, to include
Correct posture with head slightly elevated
Inhale rapidly and completely
Position of the mouthpiece (open circuit)
Exhale with maximal force
Perform manoeuvre (closed circuit method)
Have subject assume the correct posture
Attach nose clip, place mouthpiece in mouth and close lips around the mouthpiece
Inhale completely and rapidly with a pause of <1 s at TLC
Exhale maximally until no more air can be expelled while maintaining an upright posture
Repeat instructions as necessary, coaching vigorously
Repeat for a minimum of three manoeuvres; no more than eight are usually required
Check test repeatability and perform more manoeuvres as necessary
Perform manoeuvre (open circuit method)
Have subject assume the correct posture
Attach nose clip
Inhale completely and rapidly with a pause of <1 s at TLC
Place mouthpiece in mouth and close lips around the mouthpiece
Exhale maximally until no more air can be expelled while maintaining an upright posture
Repeat instructions as necessary, coaching vigorously
Repeat for a minimum of three manoeuvres; no more than eight are usually required
Check test repeatability and perform more manoeuvres as necessary
TLC: total lung capacity.

Table 2.2 European Respiratory Society /American Thoracic Society, a summary of spirometry guidelines [54].

<p>Within-manoeuvre criteria</p> <p>Individual spirometers are "acceptable" if</p> <ul style="list-style-type: none"> They are free from artefacts [3] <ul style="list-style-type: none"> Cough during the first second of exhalation Glottis closure that influences the measurement Early termination or cut-off Effort that is not maximal throughout Leak Obstructed mouthpiece They have good starts <ul style="list-style-type: none"> Extrapolated volume <5% of FVC or 0.15 L, whichever is greater They show satisfactory exhalation <ul style="list-style-type: none"> Duration of ≥ 6 s (3 s for children) or a plateau in the volume–time curve or If the subject cannot or should not continue to exhale <p>Between-manoeuvre criteria</p> <p>After three acceptable spirometers have been obtained, apply the following tests</p> <ul style="list-style-type: none"> The two largest values of FVC must be within 0.150 L of each other The two largest values of FEV₁ must be within 0.150 L of each other <p>If both of these criteria are met, the test session may be concluded</p> <p>If both of these criteria are not met, continue testing until</p> <ul style="list-style-type: none"> Both of the criteria are met with analysis of additional acceptable spirometers or A total of eight tests have been performed (optional) or The patient/subject cannot or should not continue <p>Save, as a minimum, the three satisfactory manoeuvres</p>
<p>FVC: forced vital capacity; FEV₁: forced expiratory volume in one second.</p>

2.3.2 Plethysmography

Plethysmography is used for measurement of absolute lung volumes [64]. It is completed as part of routine annual assessments at the RBH (MasterScreen™, Jaeger). For constant volume plethysmography a subject sits within a body box and the changes box pressure in relation to mouth pressure whilst completing inspiratory and expiratory efforts allow the calculation of lung volumes. The lung volume is known since the change in thoracic gas when completing an inspiratory effort against a closed shutter is proportional to the pressure rise within the box. In other words, it is assumed that no airflow is occurring during panting efforts with the shutter closed and therefore the change in alveoli pressure (inspiratory efforts are negative and slight refraction of gas occurs even without bulk gas movement)

equals the change in pressure at the mouth [48]. This is the principle of Boyles law = Pressure x Volume (at a constant temperature, the pressure exerted by a gas is inversely proportional to its volume).

One section of this thesis uses direct body plethysmography data to compare functional residual capacity (FRC) results to those generated from the MBW test (Chapter 4, section 4.4.3.1). Body plethysmography results are also used in various studies to help to identify differences between lung volume results.

2.3.3 Computed Tomography

In the MDT full CT scans were completed at screening and Follow up visits under standardised protocols on a 64-channel multi-detector scanner (Sensation 16 or 64; Siemens Medical Solutions, Erlangen, Germany). In brief, a CT comprising contiguous thin sections (1.25 mm) through the entire volume of the lungs were obtained during breath holding at full inspiration and also interspaced (1 mm sections at 10 mm increments) at end-expiration. A high-spatial-resolution algorithm was used for image reconstruction [181]. The pre and post CTs were then coded for each individual patient and time point and scored blinded by two radiographers at RBH. The final scores are a mean result from each reviewer where the presence and severity of specific CT features was recorded for each lobe using an adapted version of a semi-quantitative grading system [182] they were also compared to the Brody score [183] as completed previously by our group [147]. The maximum score is 8 or 100%. Gas Trapping, Large and small mucus plugs were used as comparisons to LCI and FEV₁%.

2.3.4 Quality of Life Questionnaire

The cystic fibrosis questionnaire revised (CFQR QoL) is a disease specific instrument measuring health related quality of life, first validated in 2005 [84]. It is a questionnaire designed to assess patient reported outcomes and has been shown to be responsive to treatment [81]. The questionnaire consists of generic and CF specific questions, designed to assess functioning during the previous two weeks. The minimal clinically important difference in the scores is 4 points (5.6 when scaled to 100), which is the difference of one category on the questionnaire (always, often, sometimes, never). This questionnaire is currently the internationally accepted field leader.

Patients completed the age- appropriate CFQR QoL (child (6-11 years or 12-13 years), parent or adolescent and adults (>14years)) at the start of the visit to avoid any influence of other

factors on the CFQR QoL scores. The results were then calculated using standardised scores. The domains available are: physical, vitality, emotion, eating, treatment burden, health perception, social, body image, role, weight, respiratory symptoms and digestive symptoms. The respiratory domain has shown to be particularly responsive in CF disease [81], this was therefore utilised in this thesis.

2.4 Multiple Breath Washout

MBW measures the volume and the concentration of a known tracer gas from its initial concentration down to the end tidal concentration (Cet) of $1/40^{\text{th}}$ (2.5%) of the starting gas concentration. Gas dilution methods only measure gas that is freely communicating within the time frame of the washout [175].

The Cet set at $1/40^{\text{th}}$ (2.5%) is historical in that this was the lowest limit of the original gas analysers used and it still utilised today [47]. There have been some suggestions to move the target i.e. with increased sensitivity of gas analysers [184] and with the use of different gases [177]. For this thesis a Cet of 2.5% will be the norm unless otherwise stated. An earlier time point has also been explored with a Cet $1/20^{\text{th}}$ (5%, LCI₅) with potential advantages of finishing the washout test quicker in turn being more manageable for young children and those with severe lung disease [185] (See Chapter 4).

Volume is derived from flow, which is measured at the mouth via a pneumotach or ultrasonic flowmeter (each described in device sections below). Expired volume of each breath is then recorded until the Cet is reached and then the cumulative expired volume (CEV) for the whole washout of tracer gas can be calculated. Gas concentration is calculated by sampling the inspired and expired gas whilst the subject is attached to the open circuit.

2.4.1 Breathing pattern

By utilising normal tidal breathing the subject is returning to a natural functional residual capacity (FRC) and the subsequent calculations i.e. lung volumes and gas mixing are normal for that patient as opposed to artificially high LCI (breaths too small) or low LCI (breaths too large). Normal tidal breathing should be appropriate for height and weight, approximately 10ml.kg^{-1} with sufficient breath size for adequate phase III slope identification and free from sighing, coughing or leaks [101].

Some laboratories utilise a fixed breathing protocol at 1L to ensure that adequate phase III slope identification can occur [186] but this is not appropriate for paediatrics to express normal gas mixing. This is because the breathing pattern is artificially high and therefore not

obtainable or repeatable [101]. It may also artificially change the FRC, which can in turn affect the output results. Historically, tidal breathing during MBW was adopted at RBH and for younger children it is deemed more appropriate. Therefore for this thesis, tidal breathing during MBW was used.

2.4.2 Deadspace

LCI is product of deadspace and ventilation distribution as these interact with the gas mixing and therefore clearance of the test gas [112]. Deadspace is accounted for in three ways, anatomical, physiological and equipment deadspace. Anatomical deadspace represent areas of the lung not contributing to ventilation i.e. the mouth, trachea and portions of the large airways. Physiological deadspace are the areas in the terminal airway where gas mix by diffusion [102]. Equipment deadspace is the portion of equipment used to send a signal to the device recording flow and gas concentrations i.e. mouthpiece, filter, pneumotach, gas pressure lines etc. Equipment deadspace correction is essential so that results are a true reflection of gas mixing in the lungs and not increased by artificially extending the airways. This is particularly important in infants and young children where deadspace is large relative to small lung volumes [187]. That being said increased LCI in infants was not associated with increased deadspace once physiological and equipment deadspace had been corrected for [188].

Current guidelines dictate that equipment deadspace should not exceed $<2 \text{ mL/kg}^{-1}$ bodyweight in children and adolescents, ideally $<1 \text{ mL/kg}^{-1}$ in infants [101]. An upper limit of 70ml is suggested for adults but there is little evidence in adults to suggest this is absolute. Previous literature has suggested that deadspace should be less than 100ml [64]. Deadspace affects LCI by essentially diluting the signal with each breath causing the LCI to appear higher when deadspace is large. With children who have smaller lungs, a larger proportion of their smaller tidal breaths will be taken up by deadspace gas so it must be kept to a minimum to ensure a true signal is being read.

2.4.3 Gases

As technological advances were made in other analysers there was a switch to using inert tracers by washing the gas in and out of the lungs. SF_6 was more abundantly used within research for measurement of LCI [74, 105], it has also been the historical device utilised at RBH within the candidates department [105, 147]. SF_6 is expensive to obtain (each bottle of air and 0.2% SF_6 gas mix is at least £250 in the UK) and specialised labs with wall brackets for

stand alone gas cylinders and gas safety monitors are required. Advances in O₂ and CO₂ analysers potentially make N₂ a more practical tracer gas for MBW testing as these gases are used to calculate the remaining N₂ with each breath within a MBW test occasion.

Gas selection may be influenced by the availability of the gas and or the equipment utilised. Each gas has different properties (Table 2.3), which may theoretically affect the final washout result. The gases utilised in this thesis are discussed in more detail below.

2.4.3.1 Tracer Gases

As described in the introductory section, SF₆ and N₂ are the main tracer gases utilised in the MBW technique. 0.2% SF₆ is utilised by the modified Innocor device. 4% SF₆ is the traditional concentration used in RMS testing, but this thesis also utilised 1% SF₆. SF₆ is measured directly by both the modified Innocor and RMS devices (see specific device sections below). As N₂ makes up the largest proportion of air there is no requirement for a washin phase, however a wait period in-between tests should be completed for the lungs to re-equilibrate with atmospheric N₂. N₂ is measured directly by the RMS device, but indirectly calculated from O₂ and CO₂ gas concentrations in the Exhalyzer D device (see Exhalyzer D section).

2.4.3.2 Washout Gases

Air (78.08% N₂, 20.95% O₂, 0.04% CO₂, 0.93% Other) has a density of 1.225 g/L. Room air is utilised in the modified Innocor technique and medical grade air is utilised in the Exhalyzer D technique. Differences between room air and medical air seem to not to have been explored from an MBW perspective. It is likely that medical air has less water vapour but as both devices use body temperature and pressure, saturated with water (BTPS) corrections (see device sections), the difference is thought to be small.

100% O₂ is utilised for N₂ washouts. 100% O₂ has a density of 1.429g/L. It is highly viscous as it has 2 spare electrons in its outer shell and being a lighter gas it has more kinetic energy thereby increasing collisions with adjacent molecules. 100% O₂ washout was also utilised in simultaneous washout section, washing out both SF₆ and N₂ at the same time.

The impact of O₂ viscosity is thought to increase with each breath during an N₂ washout as N₂ concentration is decreasing and O₂ concentration is increasing.

Table 2.3 Gas properties of the gases utilised in multiple breath washout test.

	Molar Mass (molecular weight) (Mass of a given substance divided by the amount of substance) g/mol	Density (combine molar mass and molar volume) g^L⁻¹ (typically) kg/m³ (101325 Pa, 293 K)	Dynamic Viscosity 10⁶ Pa s (10⁵ Pa, 300 K)
SF₆	146	6.5	15.9
N₂	28	1.25	17.9
Air	28.97	1.225	17.3
O₂	15.9	1.429	20.3
CO₂	44.01	1.98	16.6
He	4.003	0.1785	19.9

2.4.4 Innocor SF₆ Device

Validation of the modified Innocor device was performed by Gonem *et al.* (2014), They completed 20 combinations of simulated FRC on a lung model as well as in vivo testing on 10 HC and 14 asthma patients [146]. They found that the modified Innocor device was able to detect FRC with a measurement accuracy of 0.4%. Although MBW significantly overestimated at low and high lung volumes, the authors concluded that the modified Innocor device was accurate and repeatable at volumes within an adult age range. The modified Innocor device had also been utilised in previous studies within the PhD student's research group and was able to produce replicable results [105] showing a treatment effect [151].

2.4.4.1 Device

The modified Innocor is a gas-sampling device originally marketed for exercise testing. It has since been adapted/modified to be suitable for MBW testing [105, 113]. The modified device used here is Food and Drug Administration (FDA) approved for exercise testing but updated commercial equipment (not used in this thesis) is also approved for LCI

(<http://www.innovision.dk/Products/Innocor/Downloads.aspx>).

The modified Innocor device uses photo acoustic spectroscopy, which essentially measures the effect of absorbed energy (light) on matter by means of acoustic detection to measure CO₂, Nitric oxide (N₂O) and SF₆. A laser beam running across an unknown gas hits a known "matter" with recognised sound wavelengths and the intensity of the sound (picked up by

microphones) is proportional to the light the gas gives off, allowing the gas mixture to be quantified. Laser Diode absorption spectroscopy measures O_2 , in the same way but the Oxigraph is only sensitive to O_2 with no interference from other gases (Innocor™ Service Manual, Issue A, Revision 7, 2007). The Innocor samples gas at a rate of 100 Hertz (Hz - 100 times a second) therefore it can give an accurate reading of breath-by-breath gas concentrations.

Signal to Noise ratio compares the level of desired signal detection to the amount of background noise. The Innocor detects SF_6 at 0.2% (in the gas cylinder) at 0.001% (1/1000th) accuracy. The Innocor device has been found to have a superior signal: noise ratio compared to the mass spectrometer particularly as it measures lower concentrations of SF_6 (0.2%) and the signal stays sensitive even with falling concentrations [105]. The Innocor also has a low rise-time (the time taken for a signal to change from a low value to a high value i.e. the difference in SF_6 at inspiration and expiration, <250ms) and short delay of gas signals making it suitable for gas detection. A correction factor is incorporated for the gas delay from the Innocor however so that the final delay is smaller [105] although this does not seem to debilitate the accuracy of the measurement. In very small children, further adaptation of the device has been made i.e. removal of the oxigraph [184] since O_2 is not required to detect SF_6 measurement, but this adaptation was not made as part of this thesis.

The gas is sampled using a nafion tube that ensures equilibration of the water vapour pressure of the sampled gas with that of the ambient air. The nafion tube for the Innocor should be changed every year or when discoloured so that the absorbent properties of the nafion are not affected and therefore the gas sampled is accurate. All of the daily calibrations were performed with the correct/appropriate filter used for testing, this is to ensure that appropriate deadspace is incorporated into the system as well as ensuring no additional impact on flow will be made when using the filter. By using the same filter for calibration as for testing these issues are factored for when measurements are made; a Paediatric slim line filter (9070/01 (9070/01ZAUA)) (<http://www.gvs.com/product-family/199/928/4222>), deadspace 30ml or an adult electrostatic filter (small) (<http://www.medtronic.com/covidien/products/ventilator-filters/dar-filters>), deadspace 37ml. The adult filter was used for all testing in the MDT but the paediatric filter was used for all subsequent tests. This decision was made to be in line with the filter suggested in the Exhalyzer D SOP so that the deadspace was similar and results were more comparable.

2.4.4.2 Calibration

Further details about the calibration and testing on the modified Innocor device utilised within this thesis can be found in the updated version of the MDT MBW SOP

(http://www.mbwtraining.com/UKCFGTC_MBW_SOPsept2010.pdf).

Daily flow-gas calibrations were made so that the accuracy of the device was determined. After the warm-up period (at least 20 minutes), the operator inspired sharply and exhaled slowly 11 times. This allowed calculation of a moving average of the delay times for O₂ and CO₂. The breath-by-breath calculation contains an integration of flow and gas, and because the gas measurement is delayed through the sample line, this delay is corrected for. The delay is typically in the range 1400 to 1600ms, but can vary from instrument to instrument. The values should not vary by more than 40ms from the last flow-gas sample – if it did, then the equipment, particularly the sample lines and connections were checked and the calibration was performed again (Figure 2.1). The flow gas delay was calibrated by measuring the gas and flow change during the shift from expiration to inspiration. The calculation also takes into account the dead space volume from the inlet out to ambient air. The flowmeter that originally came with the Innocor device has a large deadspace and is inappropriate for MBW testing. Therefore a Hans Rudolph pneumotach – 4700A – 4719series flow range 0-100 L/min, dead space 18ml with a hole drilled in for the nafion sample line was used (Figure 2.1). The flowmeter contains a screen/mesh with a pressure output on each side of the screen. Pneumotachograph principle works under Ohm's Law where $\text{Flow} = \text{pressure difference} / \text{resistance}$. Therefore, with a known resistance of the instrument, measurement of the pressure drop across the pathway, then volume can be calculated by the flow signal against time.

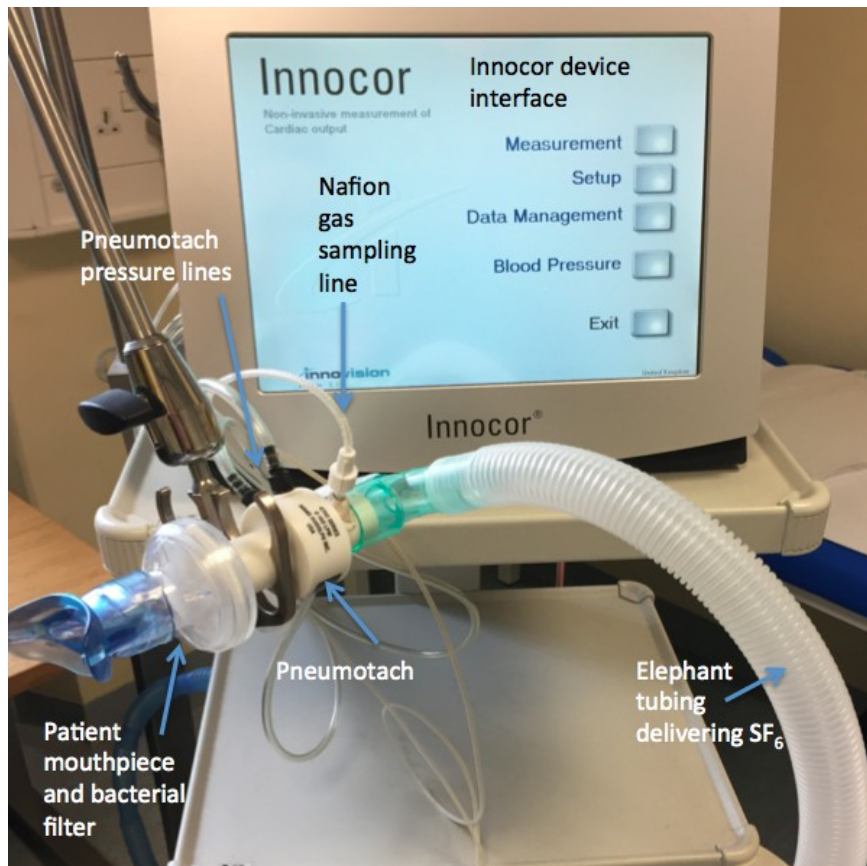


Figure 2.1 Innocor device set up indicating the patient interface and gas delivery method.

The patient attaches with the mouthpiece, there is a bacterial filter in place. The Hans Rudolph pneumotach measures flow and a nafion sample line removes a gas sample to record gas concentration. The washout of tracer gas is displayed on the Innocor screen.

Daily flow-meter calibrations were made so that the accuracy ($\pm 2\%$) of the volume being measured (3L) was checked. The operator pulled (inspiration) and pushed (expiration) the 3L syringe at low, medium and high flow volumes and providing the reading on the screen did not vary by more than $\pm 2\%$ from the 3L being measured then the calibration would be accepted. If it did vary, then the equipment i.e. sample lines, syringe and pneumotach would be checked and calibration performed again. A higher-level calibration called linearisation could also be performed. Linearisation is where a bank of varying flows for a given volume are set within the system as a reference for when the flow is recorded. This bank of data gain factors are values transforming the presumed linear flow to the actual more precise measurement where flow is not linear since it assumes a constant conductance over the entire flow range. It also interprets pressure changes in flowmeter in relation to changes in equipment over a range of flows. Linearisation was performed when the daily calibration checks failed.

2.4.4.3 Testing

The patient was sat comfortably, upright and with suitable distraction techniques in place (appropriate calming DVD). The testing module on the Innocor interface was started to ensure that flow measurement was zero before beginning. The whole mouthpiece was inserted into the mouth “like a snorkel” and nose clip in place. The patient was then instructed to breathe normally “relaxed and even breaths”.

The elephant tubing delivering the SF₆ to the patient was connected and the SF₆ bottle turned on to commence the washin phase. Wash-in was complete when the difference between the maximum and minimum SF₆ concentrations seen on the online display was 0.003% or less for three successive breaths (within 2% from SF₆ in the flow pass).

The SF₆ was then disconnected on expiration, the T piece and elephant tubing removed as quickly as possible and the SF₆ bottle turned off, this meant the patient was then breathing room air.

Washout of SF₆ was monitored on screen until the concentration had depleted to 0.003% (a cautious 1/40th of the starting concentration). A total of three technically adequate tests were ideally collected, although two were considered sufficient if breathing pattern was very similar and few testing events (swallows and pauses for example) occurred. If there was a leak during the washout then this test was discarded. Figure 2.2 shows the set up of the modified Innocor device.

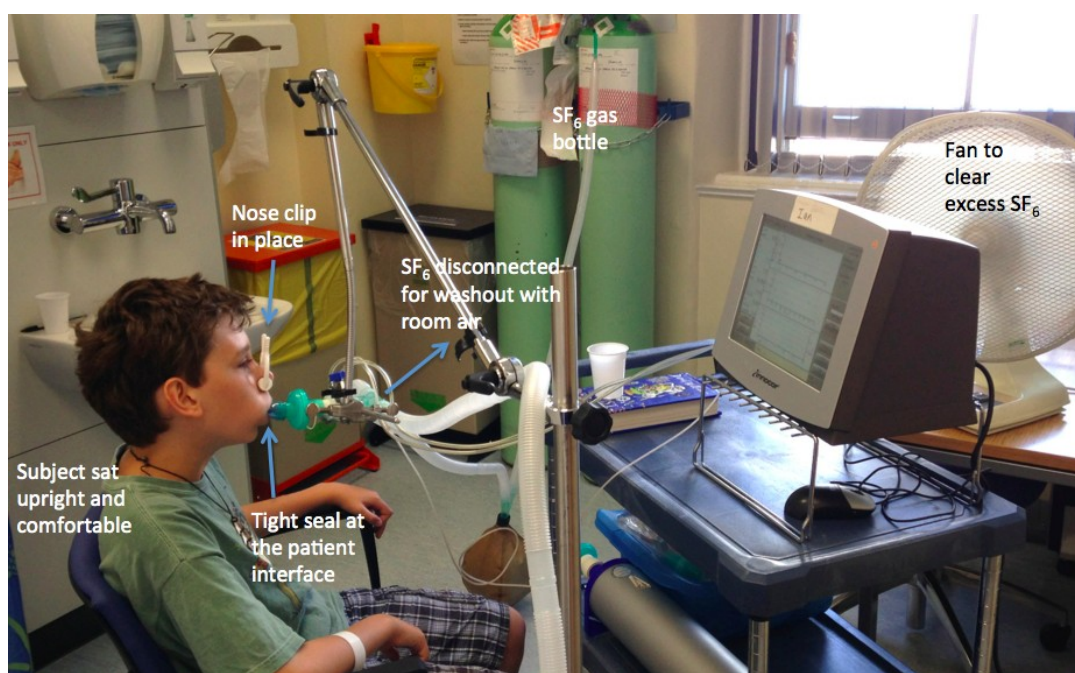


Figure 2.2 Photograph of the modified Innocor device set up. Reproduced with permission of child and parent.

2.4.4.4 Analysis

An .ino file can be exported from the Innocor device and this can in turn be uploaded onto suitable offline software for analysis of the traces. For this thesis IGOR Pro software was used, with an add in program called Simple Washout, designed and created by Dr Nicholas Bell as part of his MD within the GTC (Figure 2.3).

The .ino file must be converted into a txt file to be uploaded to the program. But first the patient ID, height and CO₂ Flow Gas Delay (FGD) were added to the yellow tab on the left of the screen. There are then capabilities to trim the washout test but essentially by pressing “Discon” the software finds the disconnection point (switch from tracer gas to breathing room air). The software is fairly automatic in that recognition of breaths and “breathlets” occur but the user must view each breath in turn to check for leaks and unusual breathing pattern. The selection of the end tidal concentration of each breath (Cet) should be accurate, re-inspired SF₆ should be low and the target Cet was reached. All traces (2 as a minimum) were uploaded and checked then an average of all 3 trials was taken as the final answer.

The main considerations for an appropriate test include that minimal difference between traces for tidal volume, FRC and LCI variability were low, and there were no obvious leaks and or variations in the traces. These factors are all highlighted in the consensus statement [101], the MDT MBW SOP (CFGTC SOP 305 MDT MBW using the Innocor v01 (a previous version to a procedure that can be found here:

http://www.mbwtraining.com/UKCFGTC_MBW_SOPsept2010.pdf and Table 5.3 in MDT Chapter 5. The simple washout software is available from Dr Nick Bell, 2010, GTC, Bristol Royal Infirmary University Hospitals Bristol, UK (See online supplement 1 produced by Hannon *et al.* 2014 [167]) it has also been further adapted and developed by Dr Alex Horsley (Institute of Inflammation and repair, Education and Research centre, University of Manchester, University Hospital of South Manchester, UK, <https://www.lci-search.com/>).

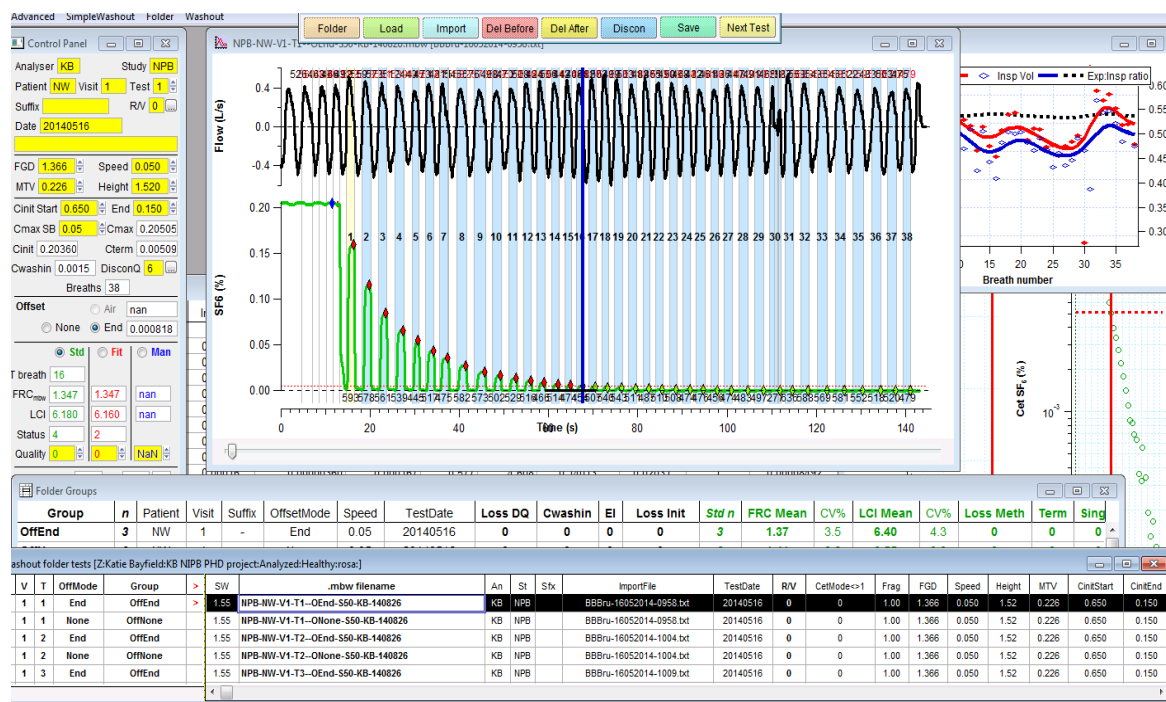


Figure 2.3 Screenshot of the Simple Washout software used for analysis of MBW traces from the modified Innocor device. The main window shows the SF₆ washout decay (green) and the other panels display the numerical results per breath and averaged from three MBW traces. The demographics and flow gas delay are entered into the panel on the left. Software produced by Dr Nick Bell (GTC).

End offset selection was used for the MDT. An offset correction takes into account any bias in the room from accumulation of SF₆ from previous tests. As SF₆ is a heavy gas, it will tend to “sink” and rest within a room if it is not appropriately ventilated – therefore when testing, a fan, an open door, air conditioning and or a long outlet/exhaust were used to minimise SF₆ accumulation in the room. The offset correction essentially adapts the whole trace so that SF₆ bias is incorporated into the analysis, i.e. setting a true zero as the ambient air used to washout the SF₆ should not contain SF₆.

2.4.5 Exhalyzer D N₂ Device

A more thorough description of the Exhalyzer D system can be found on the Ecomedics website (<http://www.ecomedics.com/index.php?id=29>) and service manual. The recommended SOP (Standard Operating Procedure; Multiple breath nitrogen washout, Jensen *et al.* 2013, version 1) can be obtained by email request from the European Cystic Fibrosis Clinical Trials Network and here: http://www.mbwtraining.com/ECFS_MBW_SOP.pdf.

Singer *et al.* (2012) completed a validation study of the Exhalyzer D. An *in vitro* study using a plastic double chamber lung model was completed as well as *in vivo* testing on 32 healthy adults. The lung model generated 145 FRC's at different breathing patterns and in 500-4174 ml, 98% of FRC's were within 5%, this is within acceptability criteria for FRC so it was concluded that the Exhalyzer D reliably measures lung volumes. FRC and LCI were calculated after MBW tests in triplicate were performed on healthy adults. FRC was more variable than the lung model but LCI was reproducible with only a 3.2% coefficient of variation [119].

2.4.5.1 Device

The Exhalyzer D contains an ultrasonic flow measuring system for measuring flow, volume and molecular mass. An ultrasonic flowmeter works by calculating the velocity of a gas across a known distance to work out the flow; an ultrasonic beam is passed between two transducers (piece 3 on Figure 2.4) and the measured transit time of the gas between the pulses of the ultrasound beam give the velocity. Ultrasonic technology can be affected by sound, temperature and potentially viscosity of a gas. But the ultrasonic flowmeter is fairly robust compared to a pneumotach/mechanical flowmeter, as it contains no moving parts.

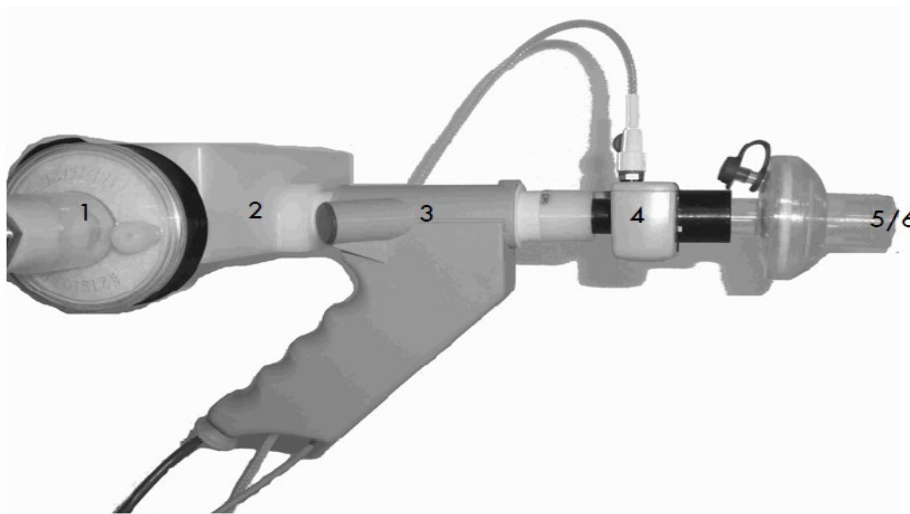


Figure 2.4 The patient interface used for the Exhalyzer D system (Ecomedics AG).

1. Bypass flow (1000ml/s) supply with filter blue, to be connected with the FRC Bypass Flow outlet (high) at the rear of the EXHALYZER D.
2. T-piece with tubing support arm/ bias flow block
3. Flow head with dead space reduce and Spirette (DSR SET 3)
4. CO2 airway adapter with connected Nafion tube
5. Bacterial filter (M30.8085)
6. Mouth piece (M30.8097)

Although the Exhalyzer D utilises a N₂ washout, it does not in fact measure N₂ directly. This is because N₂ analysers are bulky and old N₂ analyzers were slow and therefore not effective enough at sampling breath-by-breath data [47]. Instead the Exhalyzer D measures O₂ and CO₂ and derives the N₂ sample from this indirectly. Since air contains 0.934% Argon and 78.084% N₂, beta is $0.934/78.084 \approx 0.012$ therefore $N_2 = (100 - O_2 - CO_2)/(1 + \text{beta})$. O₂ is sampled through a nafion tube from side stream sampling. The patented oxygen sensor (Oxigraph – the same as the modified Innocor) uses laser diode absorption method in the infrared spectrum at a sample rate of 100Hz. CO₂ is sampled via the capnostat from the mainstream flow, the measurement utilizes advanced, self-calibrating infrared absorption technology to insure accuracy and eliminate the need for routine user calibration. The raw data from the Exhalyzer D can either be an A file which is the raw data sampled at 100Hz, B file is synchronised and C file with changeable sampling frequency. An .spx file exports the whole patient file and can be viewed and analysed within the Spiroware software.

2.4.5.2 Calibration

Further details about the calibration and testing on the Exhalyzer D system utilised within this thesis can be found in the SOP MBNW Exhalyzer D, Jensen *et al.* 2013 (http://www.mbwtraining.com/ECFS_MBW_SOP.pdf). The daily routine calibration that takes place has 5 parts: a) system settings, b) environmental settings, c) flow calibration, d) channel calibration and e) signal synchronisation.

System settings were checked to ensure the correct settings were in place for performing the test. These include, but are not exclusive to, turning off the automatic start and stop function, ensuring that the raw data will be saved in the correct format and place on the computer and checking that the correct deadspace values are in place. Automatic start and stop of test was deselected so that appropriate breathing pattern was established and therefore FRC calculation should be accurate. Of note, discrepancies and changing of default system settings occurred when different Windows log in information was used, therefore some raw data for this thesis was lost but the whole patient file retained, there did not seem to be answer to stop the error from occurring.

Environmental settings have been shown to be a key component of the system calibration in that even a small change can account for a significant impact on the final calculated results [189]. Environmental settings included checking a weather station for correct temperature,

barometric pressure and humidity of the room, the portable device was kept in the same room as the Exhalyzer D as the internal probes on the ExD are sensitive and change from floor to floor and room to room. This can be seen from a recent article where incorrect environmental settings affected the final results [189]. The Exhalyzer D manual states that if BTPS calculations are out of the valid range, then significant volume drift may occur. When indirect N_2 measurements are made from O_2 and CO_2 , the accuracy, drift and linearity of O_2 and CO_2 should meet the same performance characteristics for an MBW device [64].

For *flow calibration*, using a 1000ml syringe the gas detection pieces (nafion tube and CO_2 warmer) were removed from the system, this was because they sample some of the negative flow from the nafion tube and therefore the volume may have been affected and or the heating of the gas particles might cause deviations in the volume measured. The nafion tube access point was closed with a luer lock plug and then replicable inspirations and expirations were repeated 10 times on the flow calibration screen. The paediatric filter was in place before completing the flow calibration to ensure the impact of this was taken into consideration for the subsequent breath calculation. The flow reached ~1000ml with each pull and push of the syringe so that accuracy over the whole flow range was achieved. Three factors were considered when completing flow calibration: 1. The equipment was set up correctly i.e. there were no gaps and the Dead Space Reducer (DSR), Spirette and flow head were aligned etc. 2. The technique was appropriate for flow replication i.e. smooth with no pauses and reached the target range for all breaths. 3. The numbers were repeatable i.e. within an acceptable range (<2.5% deviation from the last testing occasion). If any of these points were not satisfied then calibration was repeated.

Channel calibration was performed after flow calibration using a set 3 DSR piece. The operator connected the nafion and CO_2 warmer as well as a gas plug to redirect the gas flow to the sampling probe from the bias flow block. Once the operator pressed 'start' the gas valves first released medical air into the system, then 100% O_2 . The operator ensured that the switch of gases was rapid and at a 90° deflection thus ensuring that the internal valve mechanism was working correctly. By checking the numbers i.e. O_2 around 21% in medical air and around 100% ($\pm 2\%$) with 100% O_2 the operator then saved the calibration. CO_2 is not displayed on screen during channel calibration; this is because at 100% O_2 a zero calibration of CO_2 was performed. It is vital that the CO_2 warmer was connected at this point, as there is not a CO_2 error message that would pop up if CO_2 were not zero i.e.

exposed to room air. The error would occur in real time sampling where it is possible to get a negative CO₂ reading thus affecting the gas synchronisation and final results. If this were to occur then calibration was performed again and then the raw A files re run with the new calibration settings.

The final part to calibration is *channel signal synchronisation*; this is where the operator performed 10 inspirations and expirations into the system at a flow rate of around 1000ml for set 3. O₂, CO₂ and molar mass (MM) were matched together i.e. for every inspiration and expiration the detection of each gas were synchronised so that zero was always detected at the same time. Synchronisation is important because N₂ is not directly measured so any misalignment of signals would be exaggerated within the N₂ trace and this can lead to vast miscalculations of the MBW outcomes. Bad synchronisation transpires as spikes and bumps on the tracings (Figure 2.8) and could be resolved by re running the raw A files with new synchronisation settings.

2.4.5.3 Testing

The subject was sat comfortably, stationary and advised to breathe calmly. The nose clip was in place. The patient demographics were added into the system then “select new test”, “multiple breath washout test” were chosen. After starting the test, gas signals appeared on screen. During the pre-phase the subject is breathing medical air; once the tidal breaths were considered stable (end expiratory lung volume did not vary for 5 breaths, flow pattern was repeatable and end tidal gas concentrations were appropriate) the N₂ washout was started. Figure 2.5 demonstrates an MBW test on the Exhalyzer D system.

By pressing, “start washout” the device automatically switched the gas to 100% O₂. There have been no safety concerns when breathing 100% O₂, particularly due to the short duration of the washout tests. The N₂ trace was monitored until it was below the 2.5% target for 5 consecutive breaths. The gas tracings were monitored closely to watch for leaks or changes in breathing pattern.

Ideally three acceptable MBW traces were completed; the operator allowed sufficient time between trials for the N₂ to re-equilibrate within the lungs [101, 190]. The Exhalyzer D system is programmed to wait 1.5 times the washout time before a new test can begin.

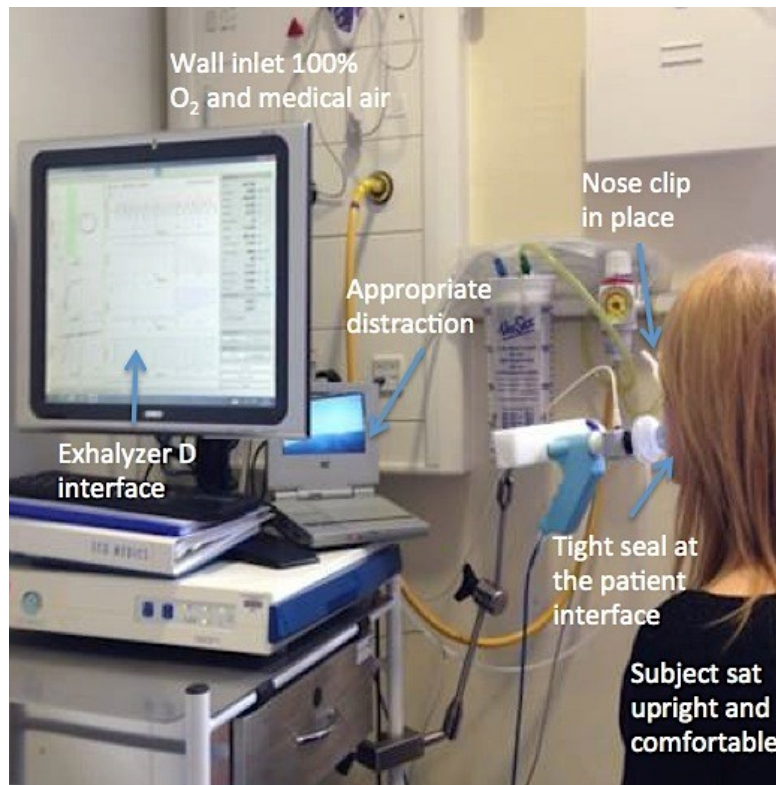


Figure 2.5 Photograph of the Exhalyzer D device set up. The patient was sat comfortably, attached to the mouthpiece and breathed wall gases (medical air then 100% oxygen) controlled electronically by the Exhalyzer D system. The operator monitored the washout trace in real time and evaluated the quality of each trial.

2.4.5.4 Analysis

Analysis on the Exhalyzer D is online i.e. calculations and algorithms are applied by the system in real time and results can be obtained at the time of testing; this software is called Spiroware (Figure 2.6 and 2.7). When overviewing the traces, results rarely change unless an individual trace was rejected. The average of at least two good quality trials is the final result. Working closely with Sickkids Hospital in Toronto, Canada, a quality control document was created [191]. In brief, a stable breathing pattern should be apparent throughout each washout (stable end expiratory lung volume, regular flow and breathing rate, Figure 2.6) and any leaks would mean the automatic rejection of a trace (unexplained rise in N₂, sudden spikes or step changes in the volume flow signal, Figure 2.8). All of the trials used in this thesis passed quality criteria specified by the core facilities.

The CTN LCI Core Facility (J. Davies *et al.* Royal Brompton Hospital, UK, for Europe), alongside the Sick Kids hospital (F. Ratjen *et al.* Toronto, Canada, for North America) and Westmead Hospital (P. Robinson *et al.* Sydney, Australia, for Australasia) have been working together to train and certify sites around the world on an aligned protocol (Jensen *et al.*

2013. Standard Operating Procedure: Multiple Breath Nitrogen Washout (Exhalyzer D)). For operators, this involves intensive training days where all machine specifics, practical testing and troubleshooting occur [168, 192]. Following this, MBW traces are submitted and analysed for quality control and all aspects of testing required for standardised, quality assured results. The Cores also perform analysis and over-reading of MBW traces for active studies using the Exhalyzer D within their trials, for example the recent Orkambi study on 6-11 year olds where LCI was used as the primary outcome [94].

Specific algorithms and calculations of key variables within the washout are embedded in the software. Recent publications suggest harmonisation of commercially available devices [165, 193] but it may be unlikely that manufacturers will ever fully align due to differences in hardware and slight differences in how to calculate certain variables. Both of the devices used here follow the specifications for equipment outlined in the MBW consensus statement [101].

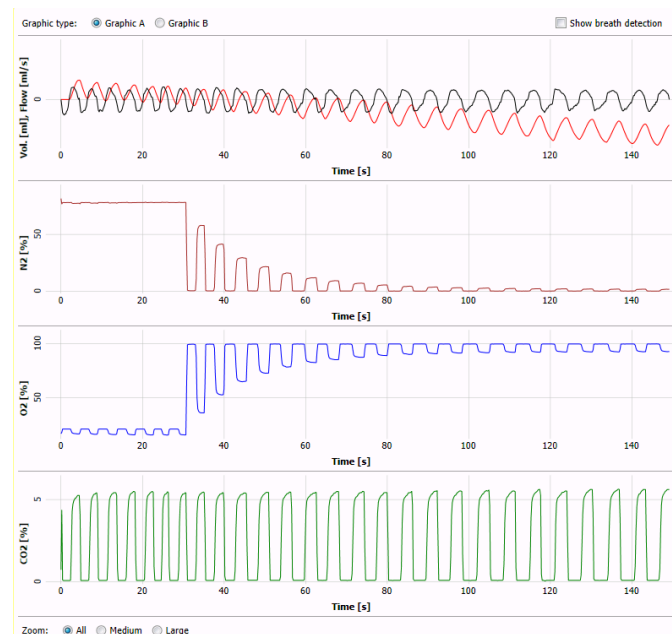


Figure 2.6 A screenshot of the sampling page from the Exhalyzer D system. The top panel shows the flow (black) and volume (red) of the trace, N₂ washout is shown in red, the O₂ washin is shown in blue and CO₂ throughout the washout is shown in green. This trace would be considered stable and acceptable.

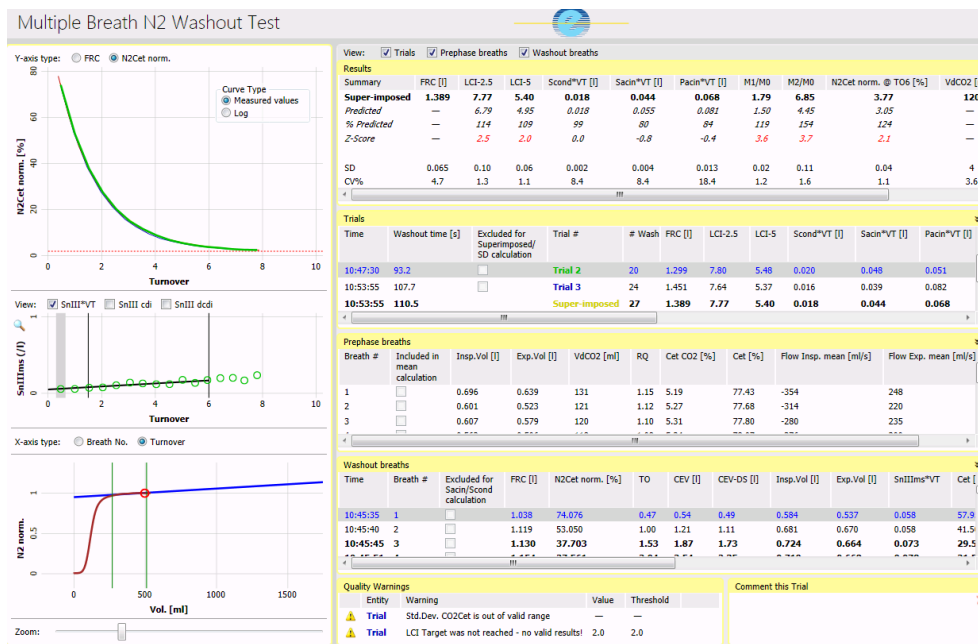


Figure 2.7 A screenshot of the analysis page from the Exhalyzer D system. The top panel shows the final results from all of the washout traces. The second panel shows the results for each individual trace. The third panel shows the pre-phase breaths for each trial. The fourth panel shows the washout breaths for each trace. Each breath can be viewed in turn, showing the N₂ washout progression and the N₂ expirogram (bottom left panel).

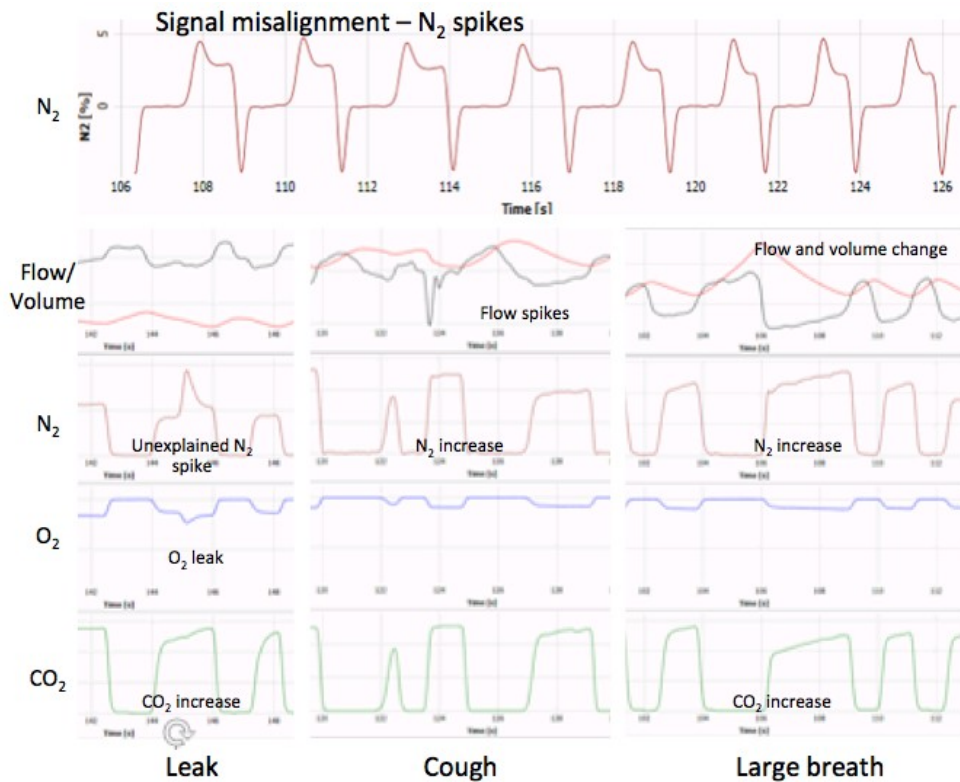


Figure 2.8 Example traces of signal misalignment, cough, leak or large breaths on the Exhalyzer D system. Sudden rises or spikes in gas traces (N_2 , O_2 and CO_2) and changes to volume/flow are typical of these multiple breath washout abnormalities.

2.4.6 Respiratory Mass Spectrometry

The respiratory mass spectrometer (RMS) has long been considered the gold standard device for simultaneous measurement of gases [101]. With low sample flow rates and short response times [194] it is ideal for fast gas analysis. It is the only analyser that fulfils all of the requirements in the recent consensus statement [101] and the only device to measure all of the gases directly. The AMIS 2000 is expensive (~£90,000) and is now out of production. Custom-built washout systems utilise the RMS as well as bespoke flow measurement devices and analysis software in specialist respiratory research centres [59, 93, 130, 159]. Hardware issues, including not being able to connect a flowmeter and a laptop to the RMS meant that this full set up was not utilised here (likely to do with a missing or faulty digital to analogue converter card). Instead just gas concentration measurement was performed using the RMS, to help to define the differences between washout tracer gases (Chapter 4).

2.4.6.1 Device

The AMIS 2000 is a scanning quadrupole mass spectrometer. Every gas has a known mass to charge ratio and the RMS can be programmed to identify specific gases based on the mass to charge ratio. Essentially when a sample is delivered into the RMS it is first ionised so that the gas and its ions become charged. These ions then pass through four charged rods (the quadrupole). The quadrupole acts as a filter, filtering the gas based on its mass to charge ratio, as the particle trajectories are influenced by the oscillating electrical fields applied onto the rods. Only those gases with known mass to ratio profiles will reach the detector where they can be identified and quantified.

2.4.6.2 Calibration

Calibration of the RMS was completed in line with an SOP developed by Dr Samantha Irving as part of her PhD, based on work by Professor Per Gustafson, which is derived from the AMIS 2000 SOP (AMIS 2000, Medical mass spectrometer system, instructions for use available upon request from Innovision, 2010, http://www.innovision.dk/Products/AMIS_2000.aspx).

As the RMS is a sensitive device, the vacuum pump should remain on at all times to remove air from the testing chamber. Full calibration at the start of each day and 2 point calibration before each individual test involved making sure that the signal in volts for each gas is as high as possible to reduce noise so that the secondary electron multiplier was optimising voltages (voltages/charge turned to mass to quantify a gas) correctly. The expensive calibration gas ($\pm 1\%$ accuracy – 4%SF₆, 4% He, 7% CO₂, 21% O₂, 64% N₂) was exposed to the sample line and the sample concentrations are referenced to known concentrations identified in the RMS. For this testing SF₆ was tuned to high priority. For the testing with 1% SF₆, a 1% calibration gas mix was not available but since 4% spanned the 1% range this was deemed adequate.

2.4.6.3 Testing

The RMS gas sample line was placed into the breathing circuit when different SF₆ and N₂ washout testing was performed (a. 4% SF₆ washout with room air, b. 4% SF₆ and N₂ washout with 100% O₂, c. N₂ washout with 100% O₂ and d. 1% SF₆ and N₂ washout with 100% O₂). For a-c a Fleisch pneumotach was used to connect the subject to the device, although it was not connected for flow recording, the small deadspace of the flowmeter was thought to be appropriate to connect the gas sample line. For d. the RMS sample line was placed in series

in between the modified Innocor and the Exhalyzer D devices that were attached using airtight plastic tubing (Figure 2.9) (Chapter 4).

Gases were sampled in real time on the RMS device; the initial starting concentration and the target end tidal gas concentration was calculated ($1/40^{\text{th}}$). All gas washouts were monitored on the RMS throughout the testing and if abnormal signals appeared (testing events such as abnormal breathing pattern or leaks) then extra testing was completed. Due to the complexity of this testing and the number of washouts for each RMS test scenario (a-d) individuals were asked to complete washouts from each only twice.

As the full washout protocol on the RMS could not be utilised, the researcher recorded time and breath number from online gas concentrations. Data was checked at the analysis phase.



Figure 2.9 The set up of simultaneous gas washout using the modified Innocor device, the Exhalyzer D and the respiratory mass spectrometer. The flowmeters were attached together and mounted so that the patient could complete a multiple breath washout of both SF₆ and N₂ using 100% O₂ as a washout gas.

2.4.6.4 Analysis

Initially, Automatic Total Pressure Correction (ATPC) was correctly set. ATPC maintains the sum of the partial pressure of the gases. The pressure change is non physiological and without correction, the gas concentrations will be affected. Checking RMS washout results were completed using the online data loading function on the RMS (Figure 2.10) ATPC was turned on. The x-axis shows time and the y-axis gas concentration; the green cursors highlight the gas concentrations at that time point and the zoom function enabled correct

identification of Cet and breath number.

The initial results (scenario a-c) were derived from online data collection in real time and using the RMS data loading function where traces were viewed on screen and gas concentration and breath number to washout recorded. Final results (gas concentration and breath number to washout) were compared between gases using t tests.

For the 1% testing (scenario d) the raw ASCII files were exported, manipulated in Excel and gas concentrations per breath derived using the simple washout software. For calculation of the end tidal gas concentration of each breath during 1% testing, the RMS gas concentration data needed to be matched with flow. The Innocor raw data file was easy to manipulate as it contains fewer parameters than the Exhalyzer D raw data file, the simple washout software is also more easily manipulated since the IGOR software is designed to handle large datasets in a variety of formats (<https://www.wavemetrics.com/products/igorpro/igorpro.htm>). First the raw data files were aligned for sample time with the Innocor raw data file. The SF₆ trace in the original Innocor file was then replaced by the SF₆ or the N₂ trace from the RMS. Data was imported into simple washout in the normal way but now the software showed the RMS gas concentrations. The gas signal per breath was calculated by the highlighting the peak of each breath, generating a mean Cet. Problems with signal synchronisation (gas from RMS and flow from the modified Innocor) meant that calculation of FRC, CEV and in turn LCI were not appropriate (Figure 2.11 and 2.12). Final results (gas concentrations and breath number to washout) were compared using t tests and plotted together to look at the decay of both SF₆ and N₂ (Chapter 4).

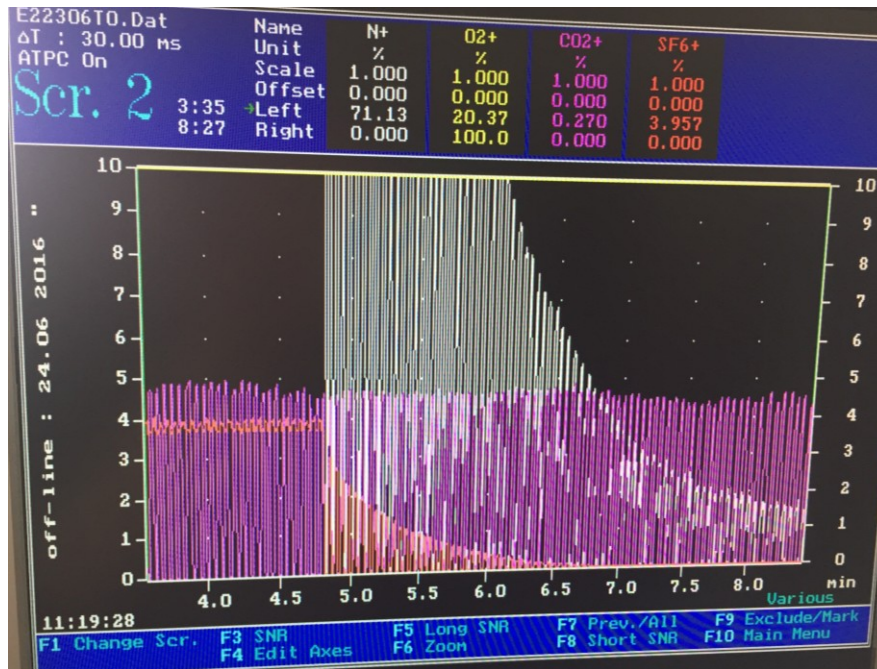


Figure 2.10 – Screenshot of the online data monitoring function on the Respiratory Mass Spectrometer device. The x-axis is time, the y-axis percentage gas concentration. This is a screenshot of a 4% SF₆ and N₂ washout with 100% O₂. CO₂ is in pink, SF₆ in red, N₂ in white and O₂ in yellow. The zoom function enabled correct identification of the end tidal concentration of gas for each breath as well as breath number to washout completion.

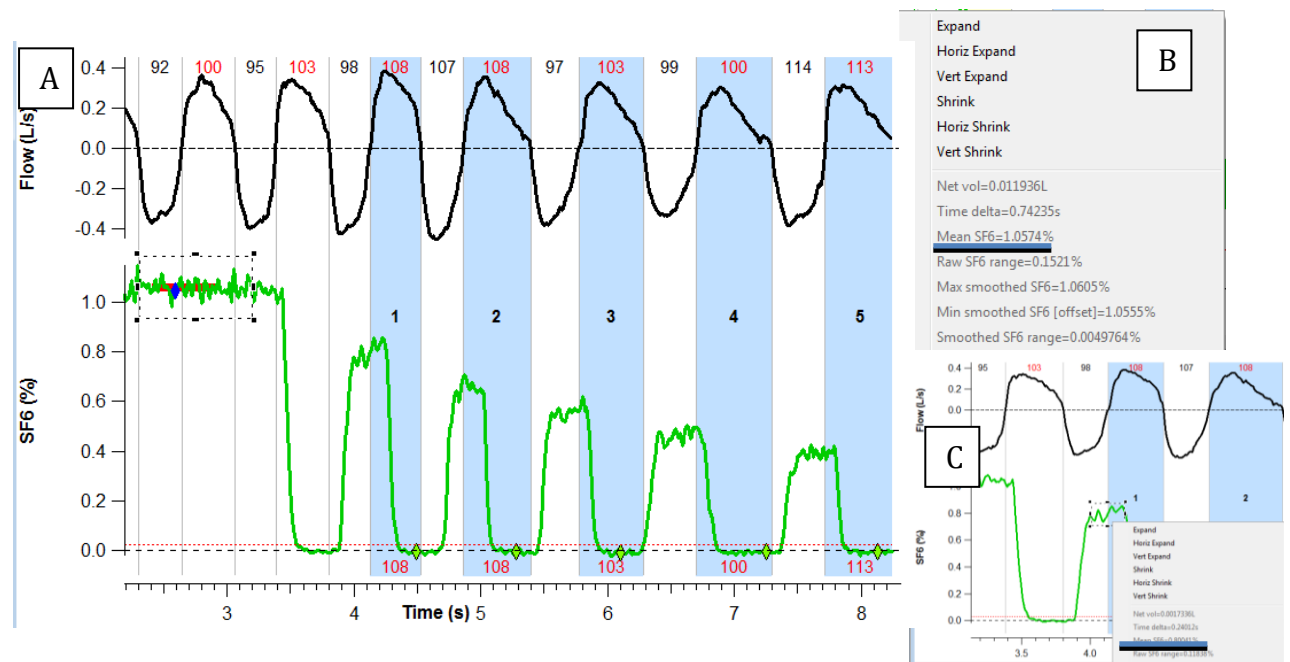


Figure 2.11 A screenshot from simple washout software loaded with the raw SF₆ data from the Respiratory Mass Spectrometer (A). The y-axis at the top shows flow (black) and the y-axis at the bottom shows SF₆ (green). The shaded blue areas represent breaths, you can see that the SF₆ (green) is misaligned with each breath (would need to be shifted to the right so that the peak of each concentration matched with flow and fell within the blue area), this is an example of the difficulties with matching different devices and software together in light of flow gas delay and different time constants. The dashed box around the SF₆ indicates where the initial starting concentration of SF₆ was taken from, by right clicking, a grey pop up box appears (B) and reveals the Mean SF₆ concentration within that area (underlined in blue). The concentration was recorded and completed for each breath until the 2.5% target was achieved (C). This was completed with all of the RMS data for SF₆ and N₂ for all subjects.

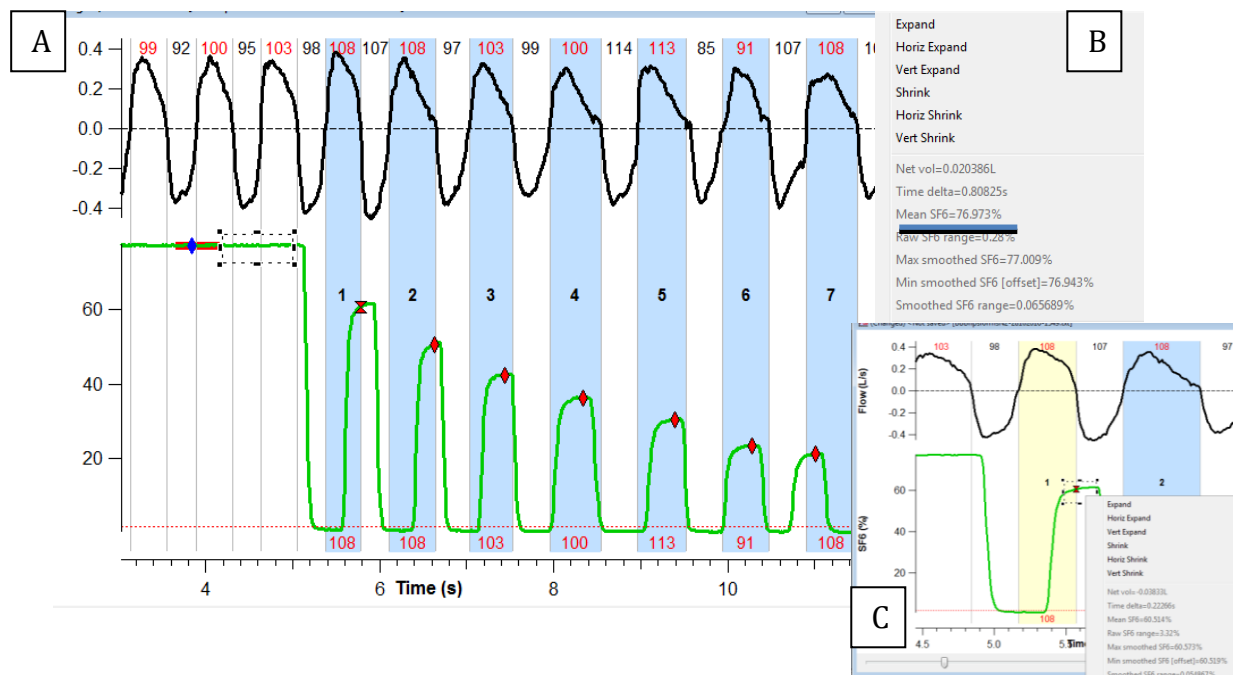


Figure 2.12 A screenshot from simple washout software loaded with the raw N₂ data from the Respiratory Mass Spectrometer (A). The y-axis at the top shows flow (black) and the y-axis at the bottom shows N₂ (green). The shaded blue areas represent breaths, you can see that the N₂ (green) is misaligned with each breath (would need to be shifted to the left so that the peak of each concentration matched with flow and fell within the blue area), this is an example of the difficulties with matching different devices and software together in light of flow gas delay and different time constants. The dashed box around the N₂ indicates where the initial starting concentration of N₂ was taken from, by right clicking, a grey pop up box appears (B) and reveals the Mean N₂ concentration within that area (underlined in blue – shown as Mean SF₆ as this is the gas programmed into the software). The concentration was recorded and completed for each breath until the 2.5% target was achieved (C). This was completed with all of the RMS data for SF₆ and N₂ for all subjects.

2.4.7 Device Specifications

Device stipulations for MBW testing are highlighted in the MBW consensus statement [101]. Errors in gas dilution method are significant if the measurement accuracy of expired volume or final gas concentration is inadequate [64]. For the devices utilised in this thesis specifications are briefly highlighted in Table 2.4. Some specifications are described in more detail below and in the device sections.

Body temperature, ambient pressure, saturated with water (BTPS)

Flow measured by a pneumotach is subject to inaccuracies from condensation of water from expiration and change in gas temperature and gas viscosity/density over the range of O₂ and N₂ mixtures [64] and influence on gas flow measurements can therefore vary by 10-15% [175]. Body temperature, ambient pressure, saturated with water (BTPS) corrections should be applied when measuring gas concentrations and flow from subjects so that results are standardised. The optimal correction would ideally be dynamic but whether the most accurate measurement of temperature and relative humidity is during inspiration or expiration is unknown [101]. All devices utilised in this thesis incorporate BTPS corrections by taking the online gas samples and adjusting the values for the final results (see Table 4).

Synchronisation

Signal alignment is essential so that the gas and volume signals match up and the final LCI result is representative of gas mixing; Figure 2.8 shows significant signal misalignment. Any delay in the machine or asynchrony could cause differences or errors in those results [195]. Alignment accuracy within 10ms or for one sample should be aimed for [101]. With increased viscosity there is an increase in time delay that causes significant synchronisation issues [196]. The SF₆ and N₂ comparison section of this thesis attempted to match gas concentration data from the RMS with flow data from both the ExD/N₂ and Inn/SF₆ devices so that FRC and LCI calculations could be made. Due to unknown flow gas delays as well as differences in gas sample calculations the signal synchronisation was deemed too complex (Appendix 1).

Software

Each device has an associated analysis software, software recommendations (online or offline) are highlighted in the recent consensus statement [101].

Global analysis software with the capacity to analyse both SF₆ and N₂ include Testpoint [59,

119, 132] and Matlab [197] although these software packages were not available for use at RBH.

The offline analysis software for the modified Innocor device was developed by Dr Nick Bell as part of his MD with the GTC (Simple washout program, IGOR Pro software, see online supplement 1 produced by Hannon *et al.* 2014 [167] or contact Dr Nick Bell (Bristol University Hospital) or Dr Alex Horsley (University of Manchester) for terms of use).

The online and offline analysis software used for the Exhalyzer D is Spiroware (<http://www.ecomedics.com/index.php?id=14>).

Viscosity corrections

Ideally a viscosity correction would be applied to all devices to account for the difference between gases and any gas delay times that may occur. As can be seen from Table 2.3, with either technique, the gas used to washout is more viscous than that being washed out, leading to accumulation of a more viscous gas mix as a tracer gas decreases. Therefore the impact of viscosity on the gas measurement may increase over the washout. However neither of the devices and their current software versions used here apply a viscosity correction.

For the Exhalyzer D, the manufacturers state that the ultrasonic flowmeter is independent of viscosity and although sidestream analysers may always be affected by viscosity i.e. increasing the flow gas delay there does not seem to be a correction for this.

For the modified Innocor technique, the flow gas delay was incorporated during analysis however the software version on the Innocor device did not have an internal correction (viscosity correction can be incorporated in software version 7 and later).

Chapter 4 describes attempts to define differences between devices. N₂ washout with 100% O₂ was completed on the modified Innocor device and calculation of an N₂ signal was attempted. Internal viscosity corrections were calculated for the modified Innocor device, however due to issues with synchronising signals the testing was aborted (Appendix 1). Significant help from Knud Pedersen at Innovision should be acknowledged here.

Table 2.4 Table of device specifications utilised within this thesis

	Modified Innocor (Innovision, Odense, Denmark)	Exhalyzer D (Ecomedics AG, Duernten, Switzerland)	AMIS 2000 Respiratory Mass Spectrometer (Innovision, Odense, Denmark)	EasyOne Pro Lab (NDD)
Hardware, version etc.	The modified Innocor uses flow meter and photo acoustic and oxigraph technology	Exhalyzer D – ultrasonic flowmeter with infra CO ₂ and oxigraph technology	AMIS 2000 – a high quality quadropole	EasyOne Pro Lab
Software, version etc.	Innovision 6.11	Spiroware 3.1.6.17312	Microsoft Disc Operating System, n/a	EasyOne Pro Lab
O₂ sensor	<p>Oxigraph O₂ sensor - The patented Oxigraph sensor uses laser diode absorption spectroscopy in the visible spectrum. Oxygen absorption is in a region of the visible spectrum (760 nm) where there is no interference or absorption by the other gases. As the oxygen concentration increases, the light intensity is attenuated, thereby identifying the concentration. The photo detector response varies linearly with the oxygen concentration.</p> <p>Principle – laser diode absorption spectroscopy – sample gas flow 120ml/min Range – 5-100% Accuracy ± 1% relative Rise time (10-90%) - <250ms Sample rate 100Hz</p>	<p>Oxigraph O₂ sensor - The fast oxygen side stream measurement module measures the oxygen respiration waveform breath-by-breath giving a qualitative indication of ventilation and the oxygen uptake.</p> <p>The patented oxygen sensor uses laser diode absorption method in the infrared spectrum.</p> <p>The lag time and sample flow is compensated by SPIROWARE®.</p> <p>Principle – side stream, laser diode absorption – sample gas flow 200ml/min Range - 2 to 100% Accuracy – 0.3% Resolutions/Linearity – 0.01% and 0.2% Rise time (10-90%) – 80ms Sample rate 100Hz</p>	<p>A high quality quadropole measures mass to charge ratio of all gases specified in the system set up. According to the gas specifications different oscillating electrical charges are applied to four cylindrical rods within the device. Because of the electrical charge applied, gas ions bounce through the rods at different trajectories. The mass to charge ratio can then be calculated for each gas. The signal is multiplied by the secondary electron multiplier (SEM), which enables detection of the gas quantity.</p> <p>Mass range 1-200 atomic mass unit Sensitivity >2x10⁻³A/mili bar Response Time 50msec Sample rate 33Hz</p>	<p>Molar mass sensor in the hand held flow unit uses ultrasonic transit time measurement (molecular weight x concentration) to calculate the molar mass of a gas.</p> <p>Molar mass analyser Accuracy <0.02g/mol Measuring range 9-50g/mol</p>

	Modified Innocor (Innovision, Odense, Denmark)	Exhalyzer D (Ecomedics AG, Duernten, Switzerland)	AMIS 2000 Respiratory Mass Spectrometer (Innovision, Odense, Denmark)	EasyOne Pro Lab (NDD)
CO₂ sensor	A photo acoustic analyser measures the effect of absorbed energy on matter by acoustic detection. A laser beam running across an unknown gas hits a known "matter" with known sound wavelengths and the intensity of the sound is proportional to the light the gas gives off, in turn revealing the gas. Principle –photo acoustic spectroscopy. Range - 0-10% Accuracy – ± 1% relative Rise time (10-90%) = <250ms Sample rate 100Hz	CO ₂ is measured directly in the patient's breathing circuit. The CO ₂ measurement module utilizes advanced, self-calibrating infrared absorption technology to insure accuracy and eliminate the need for routine user calibration. Principle – Mainstream single beam infrared, self-calibrating. Range - 0-150mmHg or 0 to 19.7%. Accuracy – 2mmHg (0-40mmHg), 5% of reading (41-70mmHg) Rise time (10-90%) = <60ms Sample rate 100Hz	See O ₂	Non-dispersive infrared (used to detect CO ₂ . A gas is passed through a sampling chamber and the CO ₂ absorbs that light at a particular frequency. A reference gas (typically N ₂ is in another chamber. By knowing the frequency and wavelength of the beam going into the machine, once it is absorbed the gas concentration can be determined. Non-dispersed means the signal is not interrupted before it is measured). Accuracy ±0.05% Concentration range 0-15%
N₂ quantification	n/a Attempts were made to indirectly calculate N ₂ from the O ₂ and CO ₂ readings. Issues with the calculation of viscosity and accounting for signal synchronisation mean the testing was aborted.	N ₂ is indirectly calculated using the following formula. $N_2 = (100 - O_2 - CO_2)/(1+\beta)$ The beta is the relative fraction of Argon/ N ₂ in air. Since air contains 0.934% Argon and 78.084% N ₂ , beta is 0.934/78.084≈0.012.	See O ₂	Combination of CO ₂ and molar mass measurement. See O ₂
SF₆ sensor	See CO ₂	n/a	See O ₂	n/a

	Modified Innocor (Innovision, Odense, Denmark)	Exhalyzer D (Ecomedics AG, Duernten, Switzerland)	AMIS 2000 Respiratory Mass Spectrometer (Innovision, Odense, Denmark)	EasyOne Pro Lab (NDD)
Analysis Software, version	Simple Washout, IGOR Pro (Wave Metrics Inc., version 6.20B3 or above)	Spiroware 3.1.6.17312	Analysis of each gas concentration was achieved using the online data loading function on the RMS (ATPC turned on, counting breaths and gas Cet from each breath). The 1% SF ₆ and N ₂ washout testing used the simple washout software to estimate the end tidal concentration from raw ASCII files from the RMS.	W breath
Signal to Noise ratio	Signal to Noise ratio compares the level of desired signal detection to the amount of background noise. So the Innocor detects SF ₆ at 0.2% (in the gas cylinder) at 0.001% (1/1000 th) accuracy. >500 @ 21% O ₂ .	O ₂ = 0.3% CO ₂ = 5% of the reading 41- 70 mmHg	Signal volts should be as high as possible (8-9 volts) in the signal optimisation calibration stage. S:N ~200 RMS auto-optimises itself as the SEM increases and decreases its signal response to different species of gas. Optimal sampling range between 0.1-4% SF ₆	QC checks signals not within 80 ppm (±3%). Exact S:N = UNK Molar mass resolution – 0.01g/mol CO ₂ analyser – resolution 0.005%
Body Temperature, ambient Pressure, Saturated with water (BTPS)	Flow is corrected for BTPS on expiration and Ambient TPS on inspiration on the device.	Correction for flow and volume signal with a fixed BTPS factor separately for inspiration and for expiration. BTPS out of range leads to Volume Drift. Correction for the O ₂ signal to BTPS condition. The CO ₂ sensor is at BTPS.	n/a	BTPS correction applied taking into account relative humidity, room temperature, atmospheric pressure or altitude.

	Modified Innocor (Innovision, Odense, Denmark)	Exhalyzer D (Ecomedics AG, Duernten, Switzerland)	AMIS 2000 Respiratory Mass Spectrometer (Innovision, Odense, Denmark)	EasyOne Pro Lab (NDD)
Viscosity	Viscosity correction is not hard coded for in version 6.11 of the Innocor device. Offline attempts were made to correct for viscosity when using the Innocor device with 100% O ₂ under guidance from Innovision but difficulties in matching signals meant that this testing was discarded.	No correction, since ultrasonic flow measurement is independent of viscosity. Possible that viscosity should be incorporated because of the sidestream analysis of gas.	Automatic Total Pressure Correction (ATPC) (Ensures the sum of the partial pressures of gases is constant (100% in line with calibration gas cylinder) despite changes to the ion source i.e. with 100% O ₂ there would be increased viscosity (gas flow changes) and total pressure of ions but the signal will account for this with ATPC on).	n/a
Flow/Volume measurement	The flowmeter contains a screen with a pressure output on each side of the screen. By measuring the pressure drop, the flow is calculated. One hertz is one cycle per second, so the sampling frequency of the Innocor is 100 samples per second. There is an automatic offset adjustment prior to each 2x5 strokes. Deviation $\pm 2\%$ relative after a gain calibration, i.e. ± 0.02 litre @ 1 litre calibration syringe and ± 0.06 litre @ 3 litre calibration syringe.	Two ultrasonic transducers mounted at different sides of the flow channel transmit ultrasonic pulses in an up a downstream direction, the measured transit times determine flow and molecular mass of the gas flow. Ultrasonic transit time detection (10ns time resolution, sampling up to 200Hz). Set 3 Deadspace reducer – flow accuracy $\pm 3\%$ or ± 5 ml/s, dead space = 20ml. If Flow calibration varies by more than 2% it should be repeated.	n/a	Ultrasonic Transit Time of molar mass, velocity and volume going through a Spirette – independent of temperature, humidity and molar mass of a gas. Accuracy $\pm 2\%$ Sample rate 400Hz Flow range ± 16 L/s
Calculation of End Tidal Concentration (Cet)	SW uses an automatic algorithm to determine SF ₆ Cet; a smoothed signal where the peak from the first peak working back from the breath-end point.	Calculates Cet from the median concentration between 90-95% of the volume of each expired breath	n/a (with this set up)	Calculates the average expired breath from 87-99% of each breath.

	Modified Innocor (Innovision, Odense, Denmark)	Exhalyzer D (Ecomedics AG, Duernten, Switzerland)	AMIS 2000 Respiratory Mass Spectrometer (Innovision, Odense, Denmark)	EasyOne Pro Lab (NDD)
Calculation of coefficient of variation	(stdev/average)*100	(stdev.p/average)*100 (resulting in a lower CV% than Innocor, this will be used so the CV% is aligned).	n/a	n/a
Signal synchronisation	Breath-by-Breath calculation contains an integration of flow and gas, because the gas measurement is delayed through the sample line. The delay is in the range 1400 to 1600ms, but can vary. A change of the inlet sample line can also change the delay. The Innocor device also “resets” the flow signal to zero every five minutes to increase accuracy. Flow and gas signals are synchronised on SW.	Asynchrony between gas and respiratory flow signals is created by transit time of gas sampling tubing and analyser response characteristics. The transit time is the time from start of exposure to the first deflection of a step change in gas concentration. The response time is the signal change from 10-90% of full signal deflection. Synchronisation values in the range of 0.500 and 0.800 s for the O ₂ and 0.03 to 0.06 s for the CO ₂ signal.	n/a (only gas signal recorded)	Ultrasonic allows determination of flow velocity and molar mass at the same time – no gas delay between signals.
Equipment Deadspace	Pneumotach 18ml. Hans Rudolph pneumotach – 4700A – 4719series flow range 0-100 L/min Adult green filter – 37ml Paediatric slim line filter – 20ml No deadspace correction incorporated into the system and was not applied in offline analysis.	Pre-Cap deadspace (volume from patient to mid-CO ₂ sensor) SET 1 [ml] (2) SET 2 [ml] (24) SET 3 [ml] (24) Post-Cap deadspace (volume from mid-CO ₂ sensor to mid flow sensor) SET 1 [ml] (3.5) SET 2 [ml] (9.5) SET 3 [ml] (22) Deadspace correction for FRC & CEV during washout.	n/a – flow not recorded	Spirette - 35ml No incorporation of deadspace into the system.
Reference ranges	[105]	[198]	[59] (Not utilised here as just gas concentrations and breath number).	[149]

2.4.8 Simultaneous Testing

Chapter 4 describes simultaneous washout of both SF₆ and N₂ using 100% O₂. The methodological steps taken to achieve the final results are shown in Appendix 1. The modified Innocor pneumotach and the Exhalyzer D ultrasonic flowmeter were attached in series (Figure 2.13). For 1% SF₆ testing using the RMS the RMS sample line was placed in between the two flow measurement devices. The same calibration methods for each device were performed as above, but the flow calibration on the modified Innocor device and Exhalyzer D were performed with the devices joined together. The devices were joined with airtight plastic tubing. The order of the devices was tested and there were no differences between results. The final setup was chosen for ease of fixing the devices in place and subjects could be sat comfortably. The total deadspace of the set up was 73.3ml. Consideration of the impact of the deadspace in inspiration and expiration is shown in Figure 2.14.

Testing was performed in much the same way as above but with slight adjustments to deliver both SF₆ in washin and 100% O₂ in washout for the washout of both gases simultaneously.

Washin – Tidal breathing was completed in washin, using a 0.2% SF₆ gas cylinder.

Switch to washout – Once equilibrium with 0.2% SF₆ in the lungs was detected by the Innocor device (normally around 2 minutes in healthy individuals) “Start WO” button on the Exhalyzer D was pressed. The gas tubing was manually swapped from the 0.2% SF₆ to the Exhalyzer D tubing. 100% O₂ (electronically switched on by the Exhalyzer D) was then breathed for washout of 0.2% SF₆ and 79% N₂.

End of washout – both gases were washed out to 2.5% of their starting concentration (SF₆ monitored on the Innocor device and N₂ on the Exhalyzer D).

Extended washout – An extended washout was performed to observe how low a gas concentration could be achieved beyond the 1/40th end tidal gas concentration. The participant continued to breathe 100% O₂ for a further 2 minutes as a reflection of the washin time of SF₆ to see if N₂ would decrease to its minimal level with similar timings as SF₆. Two further minutes of breathing 100% O₂ did not cause light-headedness in any subject.

Recorded variables - Gas concentration, breath number and time to washout were recorded rather than calculating LCI to avoid any complications related to flow volume and signal synchronisation discrepancies between the devices.

Analysis – Analysis was performed after exporting the raw breath tables from both analysis software and matching each breath with the decay of N₂ and SF₆. Deadspace and dynamic

delay time corrections were therefore not required. SF_6 and N_2 (per breath) were normalised so that they could be plotted on the same graph.

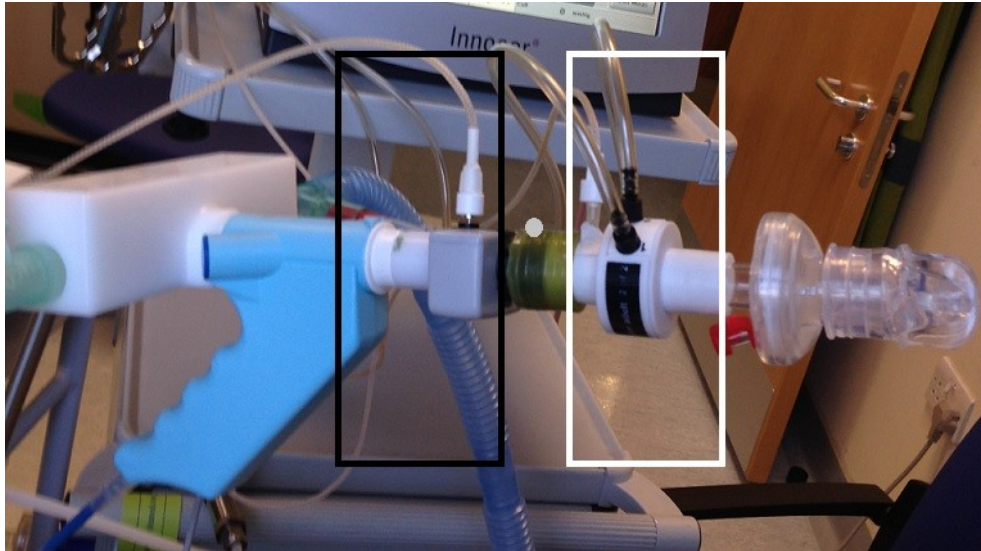


Figure 2.13 The flowmeter of the Exhalyzer D (ExD/ N_2 , black box) and modified Innocor (Inn/ SF_6 , white box) were attached in series with airtight plastic tubing; the gas concentration and breath number were then measured simultaneously. For section 4.4.3.3 a respiratory mass spectrometer (RMS) gas probe was placed between the other devices (grey dot) to read gas concentrations directly using the RMS.

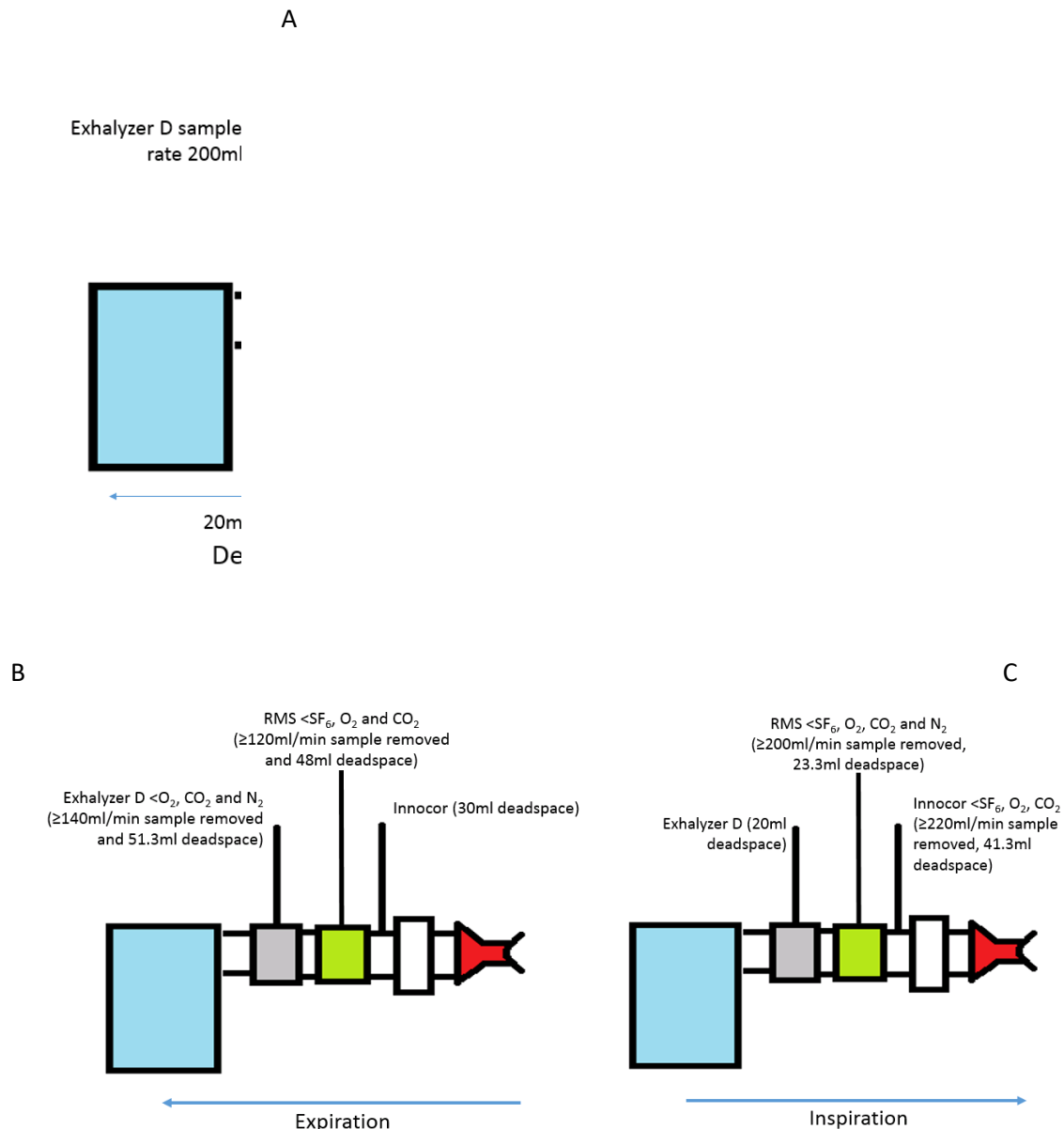


Figure 2.14 Deadspace of the simultaneous washout set up for SF_6 and N_2 washout using 100% O_2 . The Exhalyzer D and modified Innocor were attached in series by airtight plastic tubing and an RMS sample line placed in between. Graphic A shows the deadspace and sample flow rates of the set up. Graphic B shows the influence of deadspace and sample uptake during Expiration and graphic C during inspiration. Whilst the deadspace and sample uptake from each device were not included in analysis, the sum of the influence is known therefore the magnitude of potential impact they could have can be considered.

2.4.9 Statistics

Apart from Chapter 5 (the UK CF gene therapy multi-dose trial) the other sections of the thesis were not based on formal power calculations, as limited data were available using all of the techniques, specifically when exploring the differences between SF₆ and N₂. Numbers were opportunistic with attempts to keep group sizes similar.

Each section in this thesis describes specific analyses that were completed.

In general, results were analysed and Figures generated using GraphPad, Prism version 6 (GraphPad software, La Jolla, CA, USA) and Microsoft Excel (Microsoft Office 2013, Windows 7). Group differences were analysed using student paired t-tests or non-parametric Wilcoxon tests depending on prior tests of normality (D'Agostino and Pearson omnibus normality test). ANOVA analysis was used where differences in multiple groups were to be determined with *post-hoc* Bonferroni analysis to determine where differences between groups laid. Relationships between continuous and rank ordered variables were assessed using correlation coefficients and linear regression with r^2 (to determine the goodness of fit) and Chi squared used to assess patterns of change. $p < 0.05$ was considered significant throughout except when many different exploratory parameters and multiple comparisons were being compared, for which I pragmatically chose $p < 0.01$; this is highlighted in the relevant sections.

3 Chapter 3 – Method Development

3.1 Introduction

As part of the development of LCI as a surrogate outcome measure I considered four further questions to be relevant in establishing the feasibility, variability and repeatability of LCI: 1. Is there a learning effect of MBW testing? 2. Is LCI still valid when only using two replicate tests? 3. Is there a circadian rhythm to LCI? 4. Can RBH LCI reference ranges be established? All of these questions will add to the clinimetric data on LCI and potentially increase the clinical utility of the test.

3.2 Novice Patients

If a subject is new to MBW testing it may be suggested that the traditional triplicate tests will be more variable compared to when the subject has learnt the technique. This hypothesis was tested on those patients who were brand new to MBW testing and who then went on to complete a subsequent test.

3.2.1 Hypothesis, Aims and Objectives

Hypothesis 1 – Novice patients will have more variable MBW results at their first visit compared to their second

Aim 1 – Understand any learning effects with MBW testing

Objective 1 – Assess the variability of MBW testing from subjects' first and second visits

3.2.2 Methods

30 patients taking part in the gene therapy multi-dose trial (Chapter 5) were novices to LCI testing. These individuals had an Introductory (INT) visit and then went on to have a Screening (SCR) visit for entry into the trial. The LCI and LCI coefficient of variation (CV%) were compared using t tests between these two visits to assess any learning effect of MBW testing and see if the results were more variable at the first visit.

3.2.3 Results

All 30 patients had an introductory and screening visit. All patients were able to complete a valid MBW test with 3 traces available for analysis.

There were no differences in LCI or within test LCI CV% at INT or SCR (Figure 3.1) (LCI INT 10.5 ± 1.9 vs. LCI SCR 10.8 ± 2.4 , LCI CV% INT 4.4 (0.01-11.93) vs. LCI CV% SCR 5.3 (0.7-9.9)).

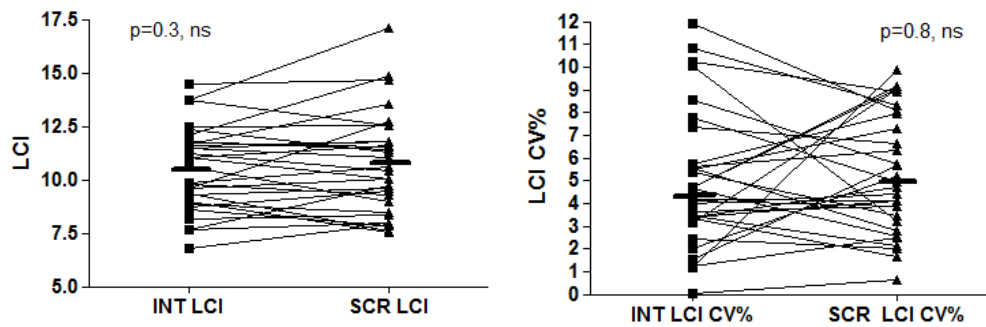


Figure 3.1 LCI and LCI CV% T test results from novice patients at their Introduction visit vs. Screening. There were no significant differences between the first novice visit and the second repeat visit, $p=ns$.

3.2.4 Summary

There were no differences between LCI or intra-test LCI CV% results between the introductory visit and screening in those patients who had both visits ($p=ns$). This implies that any learning effects, which could lead to improvements in LCI over time in novice patients, are minimal. All of the patients on the trial were above 12 years of age and this lack of learning effect would suggest they have understood how to perform the technique; the operators were able to communicate effectively and CF patients were accustomed to the requirements of lung function tests. The distraction methods used during the test may have helped. For populations similar to this group, no extra consideration appears to be needed for novice patients, as results are reproducible. This may not be the case in younger participants [133] who may not understand instructions.

3.3 Two or Three Multiple Breath Washout Tests?

Fundamental physiology testing dictates that more than one test should be completed within a testing session for the result to be regarded as valid and a true reflection of a person's lung function [65]. The MBW consensus statement confirms this philosophy, noting that three technically acceptable tests with small coefficients of variation ($<10\%$) should be completed [101]. Recent literature has explored the possibility of completing only two MBW tests instead of three as tradition dictates [167, 185], which may be particularly useful in young children where concentration and relaxed tidal breathing is difficult, but also in those with severe disease where testing is of a long duration and could be considered burdensome [133].

3.3.1 Hypothesis, Aims and Objectives

Hypothesis 1 – MBW results from 2 trials will be just as valid as MBW results from 3 trials

Aim 1 – Determine whether fewer MBW tests can be completed within a test occasion

Objective 1 – Use data from this thesis to see if there is a difference in MBW results when using 2 or 3 trials within a testing session.

3.3.2 Methods

24 HC and 18 CF patients had at least 3 MBW tests on the ExD/N₂ and 28 HC and 25 CF had at least 3 MBW tests on the Inn/SF₆ as part of data collection for this thesis. These data were explored to see if there was any loss of viable results when just the first two tests were completed i.e. outside the acceptable limits of variation as well as looking timesaving when completing just two tests.

3.3.3 Results

There were no differences between LCI from 3 or 2 tests from either the Ino and ExD. Bland Altman plots show tight limits of agreement for both devices in both HC (Inn/SF₆ -0.33 to 0.26 vs. ExD/N₂ -0.3 to 0.41) and CF (Inn/SF₆ -0.6 to 0.45 vs. ExD/N₂ -0.5 to 0.67, CF Inn/SF₆ CF ExD/N₂ Figure 3.2).

LCI CV% was higher in 3 tests compared to two in both HC and CF in the Innocor but in the Exhalyzer D three tests were not more variable than two in CF. There were no differences in LCI CV% between HC and CF.

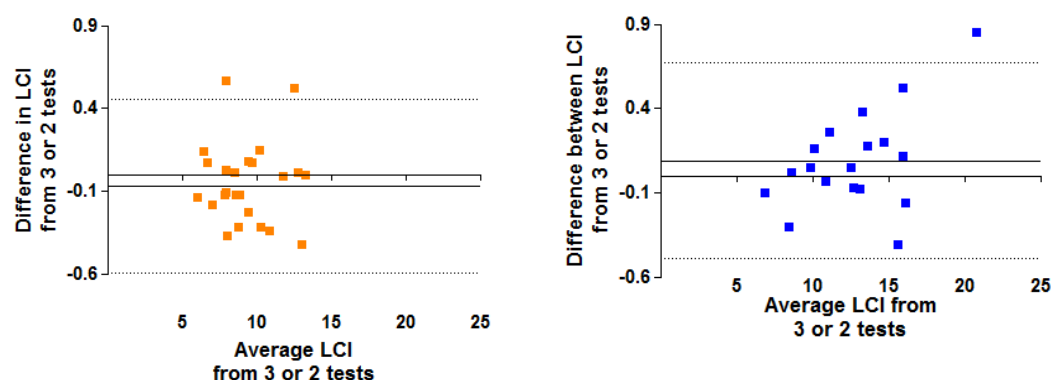


Figure 3.2 Bland Altman plots of Cystic Fibrosis LCI results from 3 or 2 tests using the Innocor SF₆ device (orange) and the Exhalyzer D device (blue).

3.3.4 Summary

3 tests introduce more variability compared to 2 tests on the Inn/SF₆ but not the ExD/N₂. Providing that the operator is sufficiently trained and understands the difficulties with MBW testing then repeatable tests should be achieved. Some individual differences may persist but having CF or not and increased disease severity do not seem to be a factor for increased within test LCI CV%.

When assessing CV% a factor to consider is whether or not online analysis forms part of the real time testing. Here, there is a difference between Inn/SF₆ (offline analysis) and ExD/N₂ (online analysis). Online analysis may already form part of the decision making process to continue to collect a third test i.e. if CV% is large from two tests vs. being within small limits of agreement for only two tests. The fact that only test occasions with three MBW tests were chosen for analysis here negates this issue. But it does imply that systems with the use of online analysis have an advantage over offline analysis in being able to fulfil the acceptability criteria with the minimum amount of tests required. This is highlighted by the total number of participants with 3 tests for this section being lower for ExD/N₂ i.e. a third test was not required as CV% was low.

As highlighted in the introduction section, subjects with increased disease severity require longer times to complete a washout. Here and in other studies, completing 2 rather than 3 tests can significantly reduce LCI testing time [166, 167, 185]. This may be a factor when considering implementation of the MBW test into the clinical setting.

For this thesis, as results here showed no LCI differences and results were still valid when generated on two devices from two or three tests, providing all quality control checks and analyses are within limits then MBW results will be grouped from subjects who performed both 2 and 3 tests. As there were variations between two tests, one test was not considered to be appropriate, yet no further analysis was completed to verify this.

3.4 Circadian Rhythm and Lung Clearance Index

Spirometry has been shown to change depending on the time of day testing was performed [199]. Medarov *et al.* 2008 highlight that even small changes can effect interpretation of results so circadian effect should be considered. Diurnal variation in MBW has not been explored. For standardisation of MBW and to limit the possible differences between devices, MBW completed in the morning and the afternoon on both the Ino and ExD system was explored.

3.4.1 Hypothesis, Aims and Objectives

Hypothesis 1 – Like FEV₁, LCI will improve in the afternoon compared to the morning

Aim 1 – Discover if LCI is affected by the circadian rhythm

Objective 1 – Complete both healthy control and cystic fibrosis MBW testing in the morning and the afternoon

3.4.2 Methods

4 HC participants completed MBW tests in the morning and afternoon on both the Ino (25 days total) and ExD (31 days total) (for every Ino time point there is matching ExD data). Testing was completed in a random order but within the same testing occasion (<1.5 hours). Morning and afternoon visits were separated by a median of 6 hours 30 minutes (1 hour 55 minutes to 8 hours 39 minutes). ExD LCIs were also performed in the morning and afternoon on 4 CF patients who were on the ward for IV antibiotics (2 have data from 2 days, 2 have data from one day). T tests assessed the differences between results from the two time points and between devices, a p value less than 0.05 was considered to be significant.

3.4.3 Results

LCI, LCI CV% and CEV do not vary diurnally on either device (Table 3.1). FRC on the Ino device was significantly higher in the morning compared to the afternoon (Ino am 2.53 ± 0.5 vs. ino pm 2.45 ± 0.5 p=0.03). This difference does not exist in the ExD (ExD am 2.95 ± 0.5 vs. ExD pm 2.9 ± 0.48 , p=ns).

LCI CV% did not differ between devices in the morning (p=ns), but in the afternoon the LCI CV% for Ino was higher than the ExD (p=0.01). When looking at paired analyses 14/25 results were derived from just two tests in the ExD, whereas 8/14 MBW tests were from 2 tests in Ino (p=0.06, ns).

Table 3.1 LCI, LCI CV%, FRC, CEV, washout breath number and V_T results for 4 Healthy Control participants completing MBW tests in the morning and the afternoon, T tests used to assess diurnal variation.

	Ino (am vs. pm) Mean $\pm$ SD or Median (range)	ExD (am vs. pm) Mean $\pm$ SD or Median (range)
LCI	6.54 $\pm$ 0.55 vs. 6.55 $\pm$ 0.6, p=ns	7.31 $\pm$ 0.37 vs. 7.37 $\pm$ 0.61, p=ns
LCI CV%	2.8 (0.08 to 7.2) vs. 2.6 (0.2 to 8.7), p=ns	1.6 (0.1 to 7.4) vs. 1.8 (0.1 to 6.2), p=ns
FRC (L)	2.53 $\pm$ 0.5 vs. 2.45 $\pm$ 0.5 p=0.03	2.95 $\pm$ 0.5 vs. 2.9 $\pm$ 0.48, p=ns
CEV (L)	15.5 $\pm$ 3.1 vs. 14.79 $\pm$ 4.2 p=ns	23.96 $\pm$ 2.9 vs. 23.94 $\pm$ 2.8 p=ns
Breaths	46.67 $\pm$ 2.2 vs. 42.5 $\pm$ 3.4 p=0.007	61.75 $\pm$ 2.5 vs. 59.11 $\pm$ 3.5 p=0.05
V_T (ml)	510.6 $\pm$ 144.1 vs. 516.6 $\pm$ 129.6 p=ns	495.6 $\pm$ 97.1 vs. 510.7 $\pm$ 115.1 p=ns

There were no differences in Cystic Fibrosis subjects for LCI, LCI CV%, FRC, CEV, washout breath number or V_T in the morning vs. the afternoon for the Exhalyzer D (Table 3.2).

Table 3.2 Cystic Fibrosis patient T test results (n=4) for Diurnal variation using the Exhalyzer D MBW device.

	am Mean $\pm$ SD or Median (range)	pm Mean $\pm$ SD or Median (range)	p value
LCI	14.43 $\pm$ 3.82	14.91 $\pm$ 4.48	p=ns
LCI CV%	1.5 (0.7-6.8)	3.15 (1.2-6.7)	p=ns
FRC (L)	2.31 (0.789-3.789)	2.41 (0.943-3.62)	p=ns
CEV (L)	45.07 (11.57-55.06)	44.41 (13.41-63.17)	p=ns
Breaths	85.3 $\pm$ 13.6	88.0 $\pm$ 10.6	p=ns
V_T (ml)	395.6 (181.1 – 779.3)	494.5 (176.3 – 807.9)	p=ns

3.4.4 Summary

There are no diurnal differences observed in HC or CF subjects for LCI, LCI CV% and CEV in either device. Further testing should be completed, as it is likely that the study is underpowered with too few numbers to fully represent any subtle diurnal changes that may occur. Further testing was not completed due to time commitments and the difficulties

faced with seeing CF patients from the hospital wards multiple times within a day. Although there is a small sample size, the differences in morning and afternoon measurements that may be exaggerated in disease are not apparent. Two studies looking at short-term variability in LCI after physiotherapy in two sets of disease patients (CF and bronchiectasis) show no change to LCI pre and post physiotherapy [158, 163]. It may be that short term, transient changes that are detected by spirometry are not evident in LCI [158]. As LCI is said to be more sensitive than spirometry this may be surprising. It could be that LCI has high sensitivity to long-term changes but is affected by all forms of heterogeneity in the short term therefore LCI has low specificity. It could also be that circadian effects only impact the large airways therefore LCI is not impacted. For these reasons timing of LCI, circadian effects and other interventions may not be important for LCI's overall readout capacity. The effect of physiotherapy on LCI from CF patients, although not significantly different, was variable between patients [163]. This highlights that understanding of LCI variability is still relatively unknown. Although circadian rhythm did not change the LCI value, continuity of testing should still be considered to ensure standardised results, relevant to individual differences.

In the Innocor device FRC was larger in the morning compared to the afternoon in HC patients but there were no differences in either device in CF. This could imply that the Innocor device is more sensitive to transient FRC changes or a false positive result related to the small number of patients. However, FRC derived from body plethysmography is found to be fairly stable throughout the day [200] providing technique is standardised [201]. It is likely that the change seen here in HC is simply due to within subject variability i.e. all values are $\pm 5/10\%$ which is within the accepted range for pulmonary tests [65]. As FRC is the denominator in the calculation of LCI and LCI did not change diurnally in this group, it is unlikely that the FRC change here is real and or the study was underpowered.

The fact that Ino LCI CV% is slightly higher in the afternoon is likely to do with small sample size and the fact that CV% may be larger when MBW tests are gathered from 3 rather than two tests. This highlights the differences between online and offline analysis where more tests are required when results cannot be seen in real time.

Interestingly the results for LCI CV% are the most variable between patients as oppose to LCI and FRC results that are fairly constant. Studies have shown that the variability of FEV₁ in asthma patients can provide additional information for asthma control than just the FEV₁ result alone [202, 203]. This has been explored in CF patients to see who is at greater risk of

lung function decline and those with greater variability showed more decline [15]. This could be a useful and an interesting outcome measure to explore within MBW.

Circadian changes in FEV₁ are well documented in healthy individuals [204]. Medarov *et al.* 2008 showed that FEV₁ was lowest at midday but increased thereafter and above FEV₁ measured in the morning. Evidence suggest that circadian effects could be reflective of changes in body temperature, oxygen consumption and carbon dioxide production, metabolic rate, exercise activity as well as hormonal changes [199]. Changes in FEV₁ in nocturnal asthma have also been attributed to increased pulmonary blood volume at night [205]. Whilst differences in MBW results were not seen here, it is likely that through thorough testing and stringent standard operating procedures that some of these issues have been avoided. For example a relaxed environment to ensure realistic tidal breathing is achieved (there were only small differences between breaths but no differences in tidal volume within devices from am to pm, Table 3.1) and ensure MBW testing is completed before spirometry or exercise tests to negate impact of forceful manoeuvres on MBW results. As differences in pulmonary blood volume have been found to change with time of day it may be that differences between SF₆ and N₂ washouts would be increased due to variable N₂ release at certain times of the day, in line with that found by Nielsen *et al.* 2013 [174]. This would need to be explored further but also exaggerates that MBW from different test gases are not interchangeable. With these things in mind it would be sensible to continue to use quality control procedures for MBW testing particularly as without large studies it is difficult to tell if circadian rhythm has an impact on MBW results.

3.5 Lung Clearance Index Z scores

LCI currently lacks established reference ranges for understanding the deviation from a healthy population. Various MBW research groups have in-house healthy control data [59, 100, 149] but currently these are not widely available. The MBW consensus statement states that software should contain normative values that include details on number of subjects, sex distribution, age range and ethnic group. [101]. The reference ranges incorporated into ExD/N₂ commercially available equipment are not fully published [198] and from the modified Innocor device, though not incorporated into analysis software, the largest reported HC data (60 HC) comes from Horsley *et al.* 2008 [105]. Utility of these reference ranges has not been assessed nor have the differences seen between SF₆ and N₂. This

section will compare predicted LCI results and z scores from the ExD/N₂ system and HC results from Horsley *et al.* 2008 to the HC data and Z scores generated as part of this thesis.

3.5.1 Hypothesis, Aims and Objectives

Hypothesis 1 – LCI Z scores from this thesis will be different between devices and from reported Z scores.

Aim 1 – Create and compare Z scores from healthy control data in this thesis

Objective 1 – Create Z scores for both the modified Innocor technique and Exhalyzer D technique

Objective 2 – Compare Z scores with existing reference ranges

3.5.2 Methods

Erika Asperges an ERASMUS student from Milan completed 80% of this section under the supervision and guidance from the researcher.

Firstly, two “Z score LCI calculators” were created from the 42 HC subjects from Chapter 4 for the Inn/SF₆ and ExD/N₂ LCI results. In Microsoft Excel, the LCI mean and standard deviation results for each device were computed to Z scores using the following formula.

$$z\text{-score} = \frac{(LCI - mean)}{standard\ deviation}$$

The predicted LCI from these Z scores could then be calculated from Z scores ranging from -5 to 5, using the formula below.

$$LCI = mean + (z\text{-score} \times standard\ deviation)$$

By creating a locked spreadsheet with these formulas incorporated, a person is able to enter an LCI result from either the Inn/SF₆ or the ExD/N₂ and have the subsequent predicted LCI and Z score result shown.

Z score results were extrapolated from the mean results reported by Houltz *et al.* 2012 [198] for ExD/N₂ (HC data from 284 people) and Horsley *et al.* 2008 [105] for Inn/SF₆ (HC data from 60 people). Using the same formula as above, the mean results (rather than raw values from RBH results) were used to extrapolate the Z score values for their datasets. The normal

distribution function in excel was then used to plot and compare the results from our HC population and the above studies.

3.5.3 Results

Using the HC data from this study, the normal range (upper and lower limits of normal set at -1.96 SD, meaning 95% of people will fall within this normal distribution) were 5.35 to 7.31 (6.33 $\pm$ 1.0 (Mean $\pm$ SD)) on the Inn/SF₆ and 5.91 to 8.66 (7.27 $\pm$ 1.39) on the ExD/N₂. Figure 3.3 shows a screen shot of the “Z score LCI calculator”, the normal range from this dataset is highlighted in green.

Innocor				Exhalizer			
LCI	z-score	z-score	LCI value	LCI	z-score	z-score	LCI value
6.45	0.230354	-5	3.828027	10.12	4.109199	-5	3.805866
		-4.5	4.078677			-4.5	4.152446
		-4	4.329327			-4	4.499026
		-3.5	4.579976			-3.5	4.845606
		-3	4.830626			-3	5.192186
Average	Standard Deviation	-2.5	5.081276	Average	Standard Deviation	-2.5	5.538766
6.334524	0.501299	-2	5.331925	7.271667	0.69316	-2	5.885346
		-1.5	5.582575			-1.5	6.231926
		-1	5.833225			-1	6.578506
		-0.5	6.083874			-0.5	6.925087
		0	6.334524			0	7.271667
		0.5	6.585173			0.5	7.618247
		1	6.835823			1	7.964827
		1.5	7.086473			1.5	8.311407
		2	7.337122			2	8.657987
		2.5	7.587772			2.5	9.004567
		3	7.838422			3	9.351147
		3.5	8.089071			3.5	9.697727
		4	8.339721			4	10.04431
		4.5	8.590371			4.5	10.39089
		5	8.84102			5	10.73747

Insert LCI value from Igor/Spiroware and press Enter. Equivalent z-scores will appear in the adjacent cell. If lower than -2 or higher than 2 it will be highlighted in red.

Figure 3.3 Screenshot of the “Z score LCI calculator” created from 42 HC participants. A range of Inn/SF₆ Z score results on the left and ExD/N₂ results on the right. The orange boxes are where an LCI value is inserted; the corresponding Z score is calculated in the blue box.

The N₂ results are higher than the SF₆ results for both HC data from this study as well as the corresponding references.

Figure 3.4 shows the LCI normal distribution curves from the HC data and reference data from the same devices but different research groups. Each dataset for N₂ and SF₆ were compared using Kolmogorov-Smirnov testing that compares the observed distribution between groups. The height of the curve represents the number of subjects within the population and the width is the range of LCI values.

For SF₆ compared to Horsley *et al.* 2008 the RBH data has a lower mean LCI and the normal distribution graph is slightly shifted to the left ($p < 0.0001$). For N₂ compared to Houltz *et al.*

2012 the HC N₂ data has a higher mean LCI, the normal distribution graph is shifted to the right and has a larger spread of data ($p < 0.0001$).

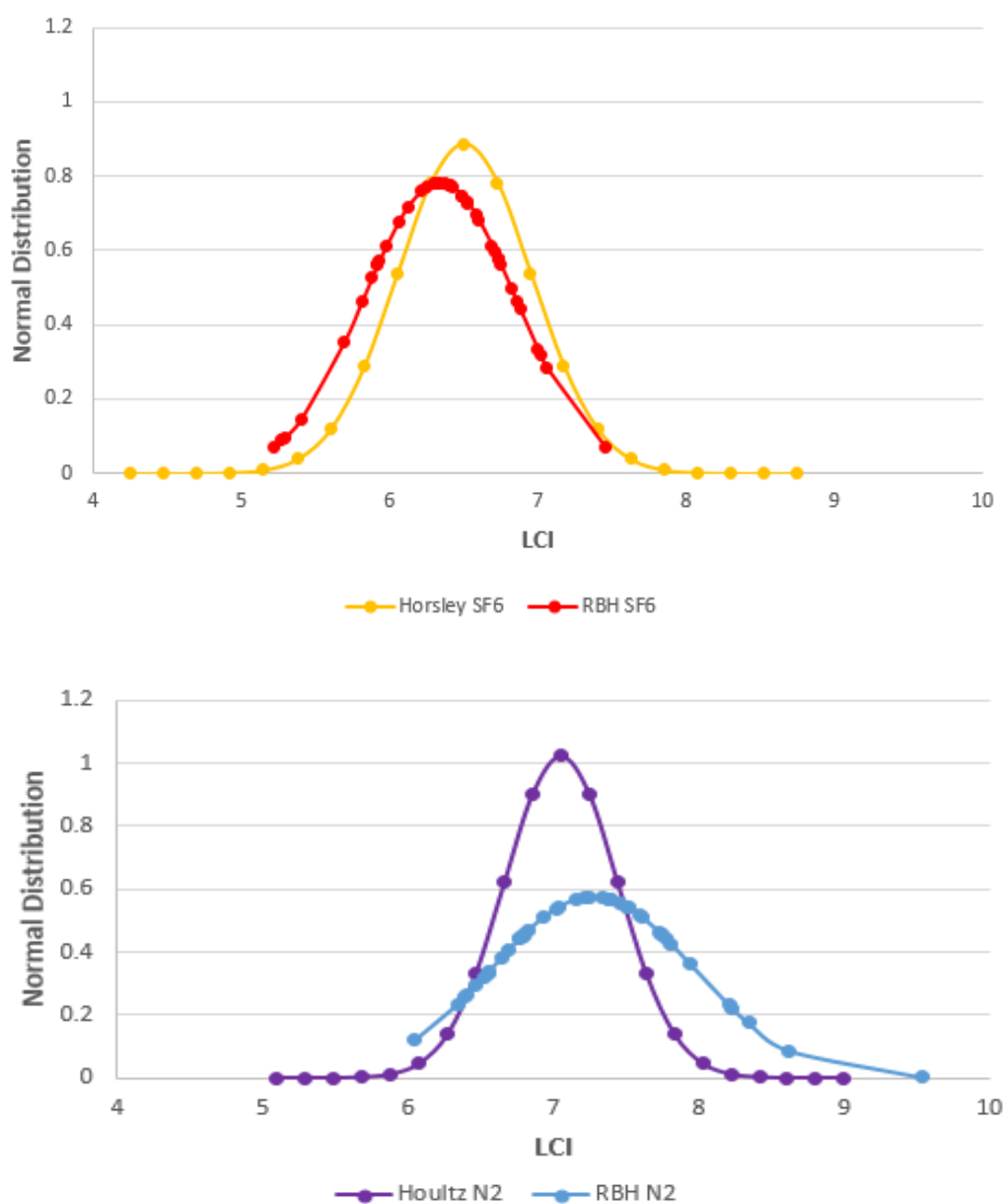


Figure 3.4 Graphs of LCI value against the normal distribution of that LCI data. The top graph shows HC data for SF₆ (red = RBH data, yellow = Horsley *et al.* 2008 [105]). The bottom graph shows HC N₂ data (blue = RBH data, purple = Houltz *et al.* 2012 [198]). Both SF₆ and N₂ slopes are significantly different from each other ($p < 0.0001$).

3.5.4 Summary

The HC data generated from this thesis are significantly different from the normal ranges generated by other researchers using the same devices. An LCI of 7.83 in N₂ could be the difference between being outside the upper limit of normal for Houltz *et al.* 2012, and fall within the normal range for the RBH reference data from this thesis. Differences like this could influence clinical decisions. For this reason, LCI Z scores were not utilised in this thesis, as the reference range at RBH was not large and robust enough to accurately relate results to the normal population.

It is stated that 120 is a valid number of participants for a reference range, with enough participants to calculate 90% confidence limits [206]. The largest reference range here is from Houltz *et al.* 2012 [198] with 284 HC. The RBH data only has HC data from 42 HC but other studies have also reported LCI Z scores from only 57 HC participants [87]. The small sample number could explain the fact that there are differences between Z scores. This emphasizes that for Z scores to be robust they need to span a large range of people. The new GLI FEV₁ references ranges are derived from a sample of 160,000 HC subjects [57], therefore it may be that LCI Z scores generated from different centers and even on specific devices are not vigorous and generalisable enough to be useful as yet.

Lum *et al.* 2015 described ethnicity as the largest influence on lung function differences and this difference persisted even if socioeconomic status was accounted for [207]. The broader range of RBH data could therefore be explained by the diversity of the country where the results were derived. According to the 2011 consensus statement, in England alone 13.8% of the whole UK population were foreign born

(<https://www.ons.gov.uk/peoplepopulationandcommunity/culturalidentity/ethnicity/articles/2011censusanalysisethnicityandreligionofthenonukbornpopulationinenglandandwales/2015-06-18>), as opposed to Sweden where foreign born nationals were 15% of the whole country population

(http://www.statistikdatabasen.scb.se/pxweb/en/ssd/START_BE_BE0101_BE0101E/UtrikesFoddaTotNK/?rxid=2af6c35f-249b-4997-ac6f-d9637612d643). That being said, the majority of the HC recruited for this study were of Caucasian origin so ethnicity may not influence the differences seen here. The ethnic background of the participants in the other studies are not available so this cannot be compared appropriately [198].

Duong *et al.* 2013 found large lung function differences across regions [208]. So although all of the sites here are within Europe, differences between places could describe the changes in Z scores. This could be due to different pollution levels between London and Sweden

(http://www.who.int/phe/health_topics/outdoorair/databases/cities/en/ and https://www.numbeo.com/pollution/compare_cities.jsp?country1=Sweden&city1=Gothenburg&country2=United+Kingdom&city2=London).

Socioeconomic class can also explain lung function differences [209]. When looking at lung function data, confounders such as age, height, body physique and socio economic status may all be considered [57, 207] and these can vary between groups. It is known that reference ranges can differ between groups [210]; caution should be taken if using supplied reference ranges from different populations as they may not be accurate and could affect clinical decisions.

LCI as a stand-alone result is a ratio of CEV and FRC and therefore it is already corrected for lung size. This means corrections for gender, height and ethnicity may not be required and the value should be applicable and comparable to all. However, LCI is influenced by age as shown by changes in LCI throughout growth [120] and increases in LCI in older healthy adults [143]. Horsley *et al.* 2008 had a slightly larger group than the HC group here (61 participants vs. 44) but had fewer LCI results from children (13 participants vs. 20). This may explain why the mean LCI from the RBH SF₆ group is lower than Horsley *et al.* 2008.

The smaller difference compared to N₂ may be a factor of the two research groups belonging to the same research consortium (GTC). So, despite the HC populations being completed 6 years apart and with no continuity of staff, the MBW methodology and standard operating procedure is likely to be very similar. This may therefore underscore the importance of standardisation of methodology particularly with the increased number of studies being completed across centers [150, 211].

Are reference ranges in LCI required?

Reference ranges provide insight as to where a result would fall when compared to a large subset of healthy control data. 95% of individuals fall within 2 standard deviations of a normal population and deviations outside of this range give information about disease. LCI has an inverse relationship to spirometry i.e. a high value in spirometry is good and any loss in percentage is associated with deterioration whereas in LCI a low value is good and deterioration is shown by increases in LCI. LCI reference ranges may therefore be misinterpreted if care is not taken when using these values. Whether the calculation of LCI Z scores should be reversed so that it is in line with spirometry i.e. a negative result being worse could be explored.

It has long been known that different reference ranges produce different results but

informed decisions on the choice of reference population and how it could influence interpretation is not as commonly applied [212]. Whilst the reference ranges calculated here may be a crude attempt and Z score calculations from lambda-mu-sigma methods may be more appropriate, as they correct for skewness in cross-sectional data [63], I have shown differences between devices and calculated ranges. Collection of new reference ranges for underrepresented populations for better understanding of lung function data between groups is advised [212]. Ideally, alongside standardisation attempts, realistic and widespread reference ranges in a large population of people, for LCI should be established for each device.

4 Chapter 4 - Defining and explaining differences in MBW observed with different techniques

4.1 Introduction

LCI has been used as an outcome measure in clinical trials, but different test gases have been utilised. Whilst the test gas does not seem to influence the benefit of LCI use in trials [38, 94], the different test gases limits the comparability of results and highlights the lack of standardisation in MBW methodology. This is particularly true since the cause of the difference is incompletely understood.

As described, the main device utilised within my department was the modified Innocor SF₆, with standard operating procedures, testing experience and a large study dataset this device was familiar and understanding of results and interpretation was high. When the ECFS CTN adopted the Exhalyzer D N₂ for standardisation of LCI within the network, I had a research platform to apply my knowledge of the Inn/SF₆ to understand the ExD/N₂, test the validity and the readout capacity of the device as well as compare the results.

It has previously been shown that differences between MBW and LCI results exist when comparing N₂ and SF₆ [118]. Table 4.1 shows that LCI results are different between health and disease as well as gas methodologies and equipment. The table highlights the differences in test gas, device and analysis software as these may all influence the results not only comparing within gas selection but between SF₆ and N₂. The Table does not provide information on individual differences between subjects but the highlighted differences could influence patient related variables i.e. respiratory rate or tidal volume [213] that could result in discrepancies between N₂ and SF₆ washouts.

Jensen *et al.* (2013) had 61 CF subjects and 62 HC subjects (age 3 to 18) complete two MBW tests, one on the Respiratory Mass Spectrometer (AMIS 2000, Innovision, SF₆ as the tracer gas) and one on the Exhalyzer D (Ecomedics, N₂ as a tracer gas,) [118]. They found that N₂ washout generated higher LCI and FRC results compared to SF₆ [118] (Table 4.1). They concluded that whilst the results were not interchangeable, the new commercially available N₂ system was a feasible alternative to the gold standard SF₆ methodology. The authors suggested that N₂ provides a better measure of ventilation heterogeneity than SF₆ as a) the N₂ result for FRC was closer to that obtained with body plethysmography and b) LCI with N₂ was proportional greater than SF₆ when disease severity increased and therefore may be more sensitive to increased heterogeneity than SF₆. Further comparison studies including correlations with other outcomes are needed to confirm this.

Poncin *et al.* 2016 completed another MBW comparison study but this time on two N₂

devices, the EasyOne Pro (New Diagnostic Design (NDD)) and the same device as Jensen *et al.* 2013 and that is currently standardised within the ECFS CTN (the Exhalyzer D) [165]. This group recruited 81 CF patients and 101 healthy controls who were able to complete two valid trials on both devices. Despite both devices utilising a N₂ washout, the Exhalyzer D system had significantly higher LCI and FRC results than the EasyOne Pro (Table 4.1). They attributed the differences to the measurement algorithm for CEV as well as differences in the demand valve presumably affecting breathing pattern since tidal volume was lower and minute ventilation higher on the NDD [165]. This shows support for the Exhalyzer D, as it does not change breathing dynamics, however, the fact that both comparison studies showed the Exhalyzer D results were above the comparator suggest that further investigation into this device is required.

Table 4.1 Key papers that complete LCI measurements on both healthy control and CF patients in a variety of ages, different devices, gases and utilising different software algorithms (Modified Innocor device).

Paper, Device, Test gas and analysis software	Healthy Controls			Cystic Fibrosis			Delta between HC and CF LCI
	Number	Age range	LCI (Mean (SD) unless stated)	Number	Age range	LCI (Mean (SD) unless stated)	
Horsley <i>et al.</i> 2008 [105] Inn/0.2% SF ₆ TestPoint	12 48	6-16 17-58	6.3 ± 0.5 6.7 ± 0.4	33	17-49	13.1 ± 3.8	6.4
Alton <i>et al.</i> 2015 [181] Inn/0.2% SF ₆ Simple washout software	n/a	n/a	n/a	Combined Pre-dosing 116	12-64	10.78 ± 2.45	4.35
O'Neill <i>et al.</i> 2016 [142] Ino/0.2% SF ₆ Simple washout software	30	27 ± 7 13-41 (2SD from mean)	6.5 ± 0.5	67	27 (16) -5-59 (2SD from mean)	10.3 ± 3	3.8

Key papers that complete LCI measurements on both healthy control and CF patients in a variety of ages, different devices and gases and utilising different software algorithms (Respiratory Mass Spectrometer).

Paper, Device, Test gas and analysis software	Healthy Controls			Cystic Fibrosis			Delta between HC and CF LCI
	Number	Age range	LCI (Mean (SD) unless stated)	Number	Age range	LCI (Mean (SD) unless stated)	
Gustafsson <i>et al.</i> 2003 [59] RMS/4% SF ₆ Testpoint	28	3-18	6.33 ± 0.43	43	3-18	8.33 ± 2.48	2
Aurora <i>et al.</i> 2004 [93] RMS/4% SF ₆ Testpoint	33	5.9-16.8	5.2 ± 2.8	22	6.4-16.5	6.2 ± 2.9	1
Lum <i>et al.</i> 2013 [120] RMS/4% SF ₆ Testpoint	242 255	0.05-6 6-19	7.33-6.45 (range) 6.54 ± 0.5	n/a	n/a	n/a	n/a
Jensen <i>et al.</i> 2013 [118] RMS/SF ₆ Testpoint	62	10.9 3-18	6.19 (CV% 0.05)	61	11 3-17	10.05 (CV% 0.05)	3.96
Ramsey <i>et al.</i> 2016 [69]ExD/4% SF ₆ Spiroware	n/a	n/a	Lum ref ranges	42	3months to 2 years	7.81 ± 0.8	1.27

Key papers that complete LCI measurements on both healthy control and CF patients in a variety of ages, different devices and gases and utilising different software algorithms (Exhalyzer D).

Paper, Device, Test gas and analysis software	Healthy Controls			Cystic Fibrosis			Delta between HC and CF LCI
	Number	Age range	LCI (Mean (SD) unless stated)	Number	Age range	LCI (Mean (SD) unless stated)	
Vermeulen <i>et al.</i> 2013 [87] ExD/N ₂ Spiroware	57	4.4-17.9	7.3 ± 0.5	63	5.3-18.8	10.78 ± 3.09	3.48
Jensen <i>et al.</i> 2013 [118] ExD/N ₂ Testpoint	62	10.9 3-18	6.81 (CV% 0.05)	61	11 3-17	11.29 (CV 0.05)	4.48
Poncin <i>et al.</i> 2016 [165] ExD/N ₂ Spiroware	101 (51 adults, 50 children)	8-28	6.92 ± 0.35 7.05 ± 0.55	81 (34 adults, 47 children)	9-33	13.73 ± 3.63 9.76 ± 2.26	6.81 2.71
Ramsey <i>et al.</i> 2016 [69] ExD/N ₂ Spiroware	72	3-15	6.7 5.7-7.7 (LLN and ULN)	39 38	3-6 7-16	8.17 ± 1.2 8.35 ± 1.8	1.47 1.65
Anagnostopoulou <i>et al.</i> 2016 [214] ExD/N ₂ Spiroware	136	10 2.8-18.1	7.2 ± 0.5	n/a	n/a	n/a	n/a
Gustafsson <i>et al.</i> 2014 [129] ExD/N ₂ Testpoint	74	23 (7.3) 8.4-37.6 (2SD from mean)	6.67 ± 0.35	37	23 (7.7) 7.6-38.4 (2SD from mean)	12.16 ± 3.33	5.49

Key papers that complete LCI measurements on both healthy control and CF patients in a variety of ages, different devices and gases and utilising different software algorithms (EasyOne Pro (NDD) and custom equipment).

Paper, Device, Test gas and analysis software	Healthy Controls			Cystic Fibrosis			Delta between HC and CF LCI
	Number	Age range	LCI (Mean (SD) unless stated)	Number	Age range	LCI (Mean (SD) unless stated)	
Fuchs <i>et al.</i> 2012 [150] NDD/4% SF ₆ W breath	102	12.5 (5.1-19.5)	6.3 ± 0.19	139	11.8 (4.3-23.6)	8.17 ± 1.85	1.87
Poncin <i>et al.</i> 2016 [165] NDD/N ₂ W breath	101 (51 adults, 50 children)	8-28	6.42 ± 0.54 6.78 ± 0.63	81 (34 adults and 47 children)	9-33	10.01 ± 2.46 8.32 ± 1.64	3.59 1.54
Verbanck <i>et al.</i> 2013 [128] N ₂ analysis, 1L tidal breaths Custom software	60 F 60 M 60 F 60 M	25-65 25-65	6.26± 0.4 6.28 ± 0.4 5.77 ± 0.5 5.65 ± 0.5	n/a	n/a	n/a	n/a
Verbanck <i>et al.</i> 2013 [114] N ₂ analysis, 1L tidal breaths Custom software	25	27 (24-30)	6.2 ± 0.4	25	28 (24-29)	10.3 ± 2.3	4.1

N₂ derived LCI results are higher when compared to SF₆ on the same individuals [118] but the mechanism behind this is still unclear.

The four main explanations for the difference in gas washouts are:

- Individual differences between washout systems i.e. breathing pattern,
- Tissue N₂ contributing to a N₂ washout signal,
- Equipment differences,
- Inherent gas differences; tracer and washout gas.

Individual differences and natural variability exist within lung function testing, this is why repeat measures are performed [54]. As the MBW tests are performed on different devices at different times this variability could be the cause of the differences seen.

As described previously, tissue N₂ may contribute to the washout signal during a N₂ MBW test using 100% O₂ as washout gas. This could explain the differences between N₂ and SF₆, particularly as SF₆ is an entirely inert gas. Figure 4.1 illustrates the different sources of N₂ from the body that could add to the N₂ washout signal.

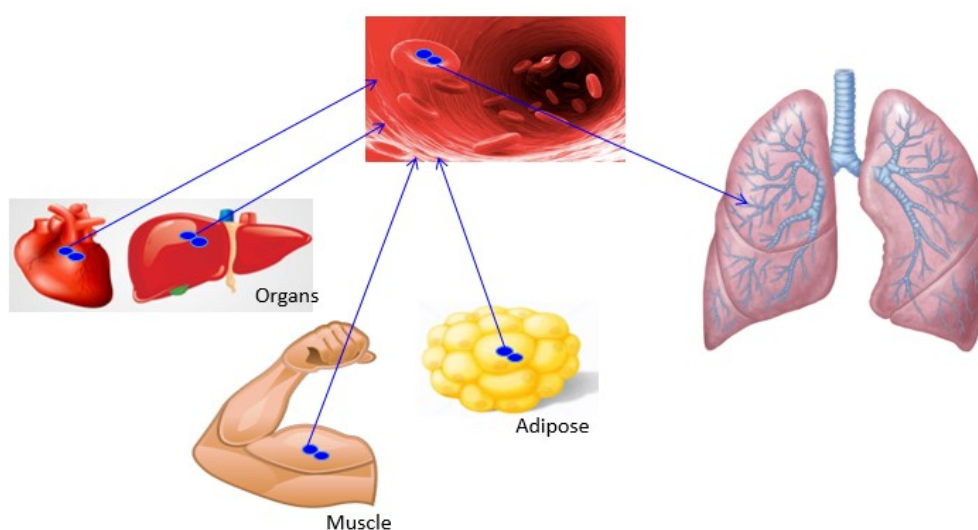


Figure 4.1 Diagrammatic examples of sources of innate N₂ that may contribute to and extend the washout signal in N₂ MBW.

Table 4.1 shows different LCI results when completing MBW testing on different pieces of equipment. All devices follow the ERS MBW consensus statement but there are software and hardware differences between devices that could contribute to the difference in LCI results.

Another reason for the discrepancies between techniques using different test gases could be the inherent properties of those gases. Gas diffusion is defined as gas molecules moving

from a high concentration to a low concentration; Fick's first law of diffusivity [49]. Each gas has its own diffusion coefficient related to gas concentration and distance. Gas is conducted down the airways to the gas-exchanging zone by convection and then molecular diffusion. There is a point where both of these processes occur – the convection-diffusion front that is located around the acinus entrance. Diffusivity is proportional to molar mass and affects the diffusion convection front [47]. The diffusion convection front is relatively stationary for each gas, being more proximal for N_2 at the terminal bronchiole [215] than SF_6 (more peripherally located for SF_6 (generation 18) [47]). The concentration gradient between the bulk flow and resident gases aims to equilibrate each gas concentration in turn [47, 49] and there are greater asymmetries deeper in the acinus for SF_6 [215].

Jensen *et al.* 2013 also describe that the resident nature of N_2 may reveal more gas trapping and gas mixing from poorly ventilated areas than SF_6 as SF_6 is an entirely inert gas that may only partially washin due to obstruction in turn leading to lower LCI values [118], this explanation is described in Figure 4.2. However as the convention diffusion front is more peripherally located for SF_6 compared to N_2 [47, 215] this explanation may only exist in severe disease.

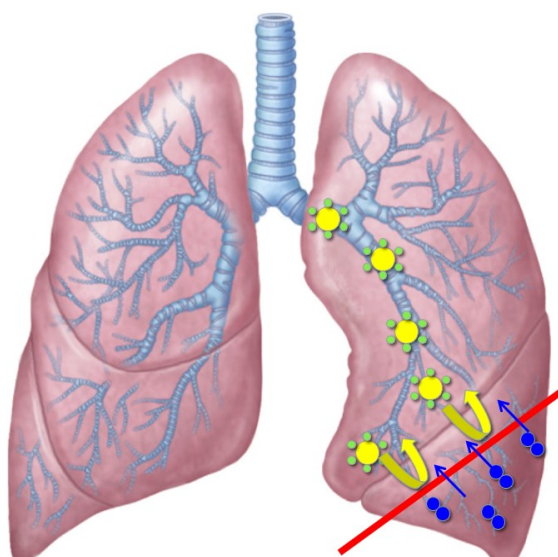


Figure 4.2 – Diagrammatic representation of the potential partial washin of SF_6 (Yellow and green) to only the larger communicating lung units. Therefore a hypothesis for LCI results being lower with SF_6 is that it fails to reach distal airways in severe disease.

Whereas N_2 (blue) is not an inert gas and may be able to give ventilation heterogeneity information from more distal, smaller lung units and potentially even where gas trapping (distal area beyond the red line) is apparent [118].

This thesis chapter will describe a series of prospective experiments designed to explore the differences in SF₆ and N₂ washout and attempt to differentiate between the four explanations offered. Figure 4.3 describes the major aims running through this Chapter and the explanations considered most likely for the differences in SF₆ and N₂ MBW results. Kent *et al.* 2013 suggest that further analysis and comparisons of this kind are required for creation of normative ranges and to define the coefficient of repeatability across ages [52]. Further understanding and interpretation of results will help the standardisation of LCI and define LCI's potential as a surrogate end point in clinical trials.

4.2 Hypothesis, Aims and Objectives

Hypothesis 1 - LCI measured using N₂ washout on the Exhalyzer D provides a robust signal of gas mixing inhomogeneity, correlating with conventional measures of airway disease severity.

Hypothesis 2 - LCI measured with N₂ will correlate with, but differ in absolute value from, LCI values obtained with SF₆. These results will not be interchangeable.

Aim 1 – Compare N₂ and SF₆ LCI results.

Objective 1 - Complete a comparison of the Exhalyzer D (technique standardised around by ECFS CTN) and modified Innocor device (established technique in our department) with both healthy control subjects and those with CF.

Objective 2 - Examine the impact of shortened washout by comparing LCI_{2.5} (1/40th starting concentration) and LCI₅ (1/20th).

Aim 2 – Examine whether the differences between devices are impacted by disease severity.

Objective 3 - Correlate FEV₁ results with LCI from both Exhalyzer D and modified Innocor.

Objective 4 – Determine which patient group LCI monitoring is most beneficial for.

Aim 3 – Understanding LCI differences between gases by questioning the following:

- a. Does FRC cause the difference between devices?
- b. Do individual patient factors between tests contribute to the difference?
- c. Do the tracer gases washout differently? Does breathing 100% O₂ impact the washout?
- d. Does using the respiratory mass spectrometer (RMS) to measure both gases eliminate equipment differences?

Objective 5 - Compare FRC results from both techniques to those obtained with body plethysmography.

Objective 6 – Complete a simultaneous washout on both the modified Innocor device and the Exhalyzer D to eradicate between test differences.

Objective 7 – Complete an extended washout to compare differences between tracer gas washouts.

Objective 8 – Complete a simultaneous washout of both gases on the RMS to establish if equipment differences between the Innocor and Exhalyzer D are the main cause of the discrepancy between results.

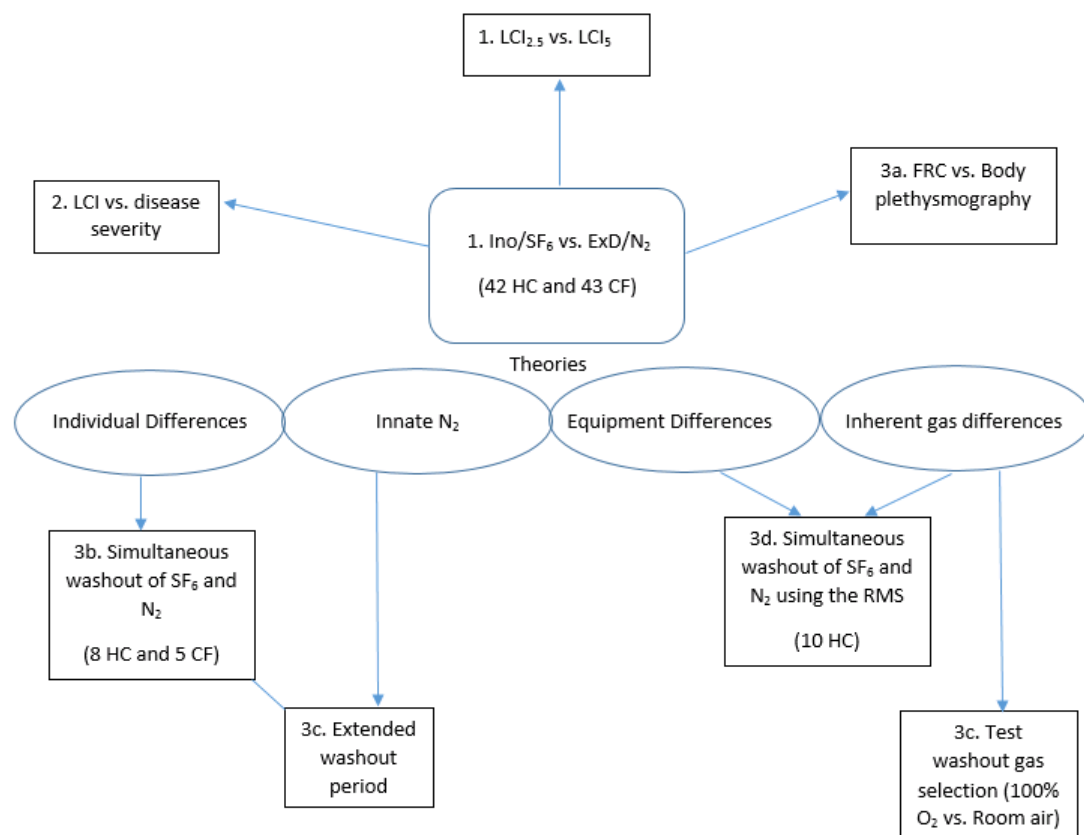


Figure 4.3 – Process diagram of the three main aims and subsections for this chapter.

4.3 Methods

Specific methodology utilised for the different sub sections is described in each results section for ease of reading and continuity with the aims.

4.4 Results

4.4.1 Compare N₂ and SF₆ (Aim 1)

Methods: CF and HC were asked to complete 3 MBW trials on two devices: the Exhalyzer D, N₂ washout (ExD/N₂, Ecomedics AG, Switzerland) and the modified Innocor™ gas analyser (Innovision, Denmark), SF₆ as a tracer gas (Inn/SF₆); detailed methods in the Methods chapter of this thesis section 2.4.3 and as previously described [105, 118]. Participants were allowed to only perform two MBW tests on each device if there were time constraints related to attempts to minimise inconvenience to the patient by recruiting during a clinic. Data shown in section 3.3 indicate that two tests are acceptable and appropriate for this thesis.

The ExD/N₂ and Inn/SF₆ were performed on the same test occasion but in an alternate order between subjects. Subjects breathed tidally throughout each washout either inhaling 0.2% SF₆ followed by room air for Inn/SF₆ or medical air followed by 100% O₂ for ExD/N₂. The washout gases were monitored until they reached the historic cut off point of 1/40th of the starting concentration (LCI_{2.5}); for the purposes of objective 2, traces were also assessed at the 1/20th cut off point (LCI₅). Real time assessments of breathing pattern, leaks or abnormalities were also made and traces excluded according to the 2013 consensus statement [101] (see methods section). To maintain a calm, relaxed and similar tidal breathing pattern between devices, the same calming DVD was played throughout testing. Healthy Control (HC) and CF MBW parameters were compared within and between groups using T Tests or Wilcoxon signed ranked tests and Bland Altman analyses. FEV₁% and LCI_{2.5} and LCI₅ results from both devices were plotted and compared using linear regression analyses. I hypothesised that this analysis would reveal where in the washout trace differences between gases were occurring.

Results: 47 CF participants and 42 HC subjects completed MBW tests on the Inn/SF₆ and ExD/N₂ on the same day. Table 4.2 shows the subject demographics. CF participants tended to be younger, shorter and weighed less than the HC but the differences were not statistically significant for any parameters.

Table 4.2 Subject demographics for CF and HC individuals who completed at least two MBW tests on the Exhalyzer D and modified Innocor device on the same day (results represented as mean $\pm$ SD unless otherwise stated). No variables were statistically different between groups.

(Inn/SF ₆ vs. ExD/N ₂)	CF (n=47)	HC (n=42)
Gender	24 (51%) Female	26 (62%) Female
Age (years) (median (range))	16.05 (5.9-63.7)	24.32 (5.7-56.1)
Height (cm) Mean $\pm$ SD	154.6 $\pm$ 52.2	158.4 $\pm$ 17.9
Weight (kg) Mean $\pm$ SD	52.2 $\pm$ 18	59.7 $\pm$ 20.6
FEV₁ (%) (Z Score) Mean $\pm$ SD	72.9 $\pm$ 16.8, (-2.29 $\pm$ 1.43)	Not Done

100% of HC participants and 93% of CF patients were able to complete at least 2 acceptable MBW tests on both machines. The MBW traces from 3 CF patients were excluded at analysis due to poor quality: 2 patients had a lack of repeatability in FRC (>25%) and LCI and 1 patient had a leak in the washout that was shown on both devices.

The intra-test coefficients of variation in LCI (CV%; median (range)) were small and not statistically different between devices (HC Ino 2.85% (0.2-11.8) vs. HC ExD 2.87% (0.1-10.9) p =ns and CF Ino 3.7% (0.02-13.02) vs. CF ExD 3.06% (0.1-7.4) p =ns) or between HC and CF subjects (p =ns).

As expected, LCI_{2.5} (mean $\pm$ SD) was significantly higher in CF than HC; this was shown with both devices (Table 4.3). An illustration of an individual washout from each device is shown on page 132.

In both CF and HC, the ExD/N₂ device gave higher LCI readings than the Inn/SF₆. In HC there was a narrow range of results between devices and narrow 95% limits of agreement (-0.38 to 2.22, Figure 4.4). However, in CF, the range of LCI values was wide, the limits of agreement were larger (0.08 to 8.92, Bland Altman plot, Figure 4.4) and the difference between devices increased with greater LCI (increased disease severity), this resulted in a significant regression line (p <0.0001, not shown on the Bland Altman plot).

LCI₅ was also higher in CF compared to HC and in the CF patients was higher in ExD/N₂ compared to Inn/SF₆ (Table 4.3).

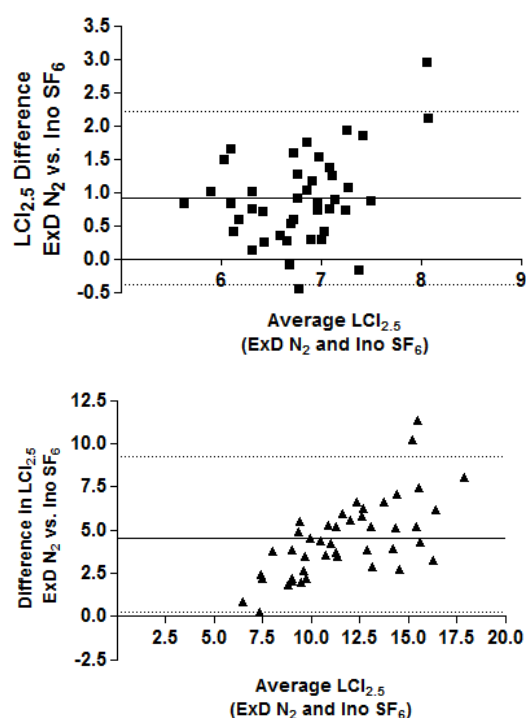


Figure 4.4 Bland Altman plots of the LCI difference between ExD/N₂ and Inn/SF₆ in HC participants on the left and CF patients on the right. Note the differences in axes scales, smaller in HC. The solid line is the mean difference between devices and the dotted lines represent the limits of agreement (-0.38 to 2.22 in HC and 0.08 to 8.92 in CF). In both cases the variance depends strongly on the magnitude of the difference.

Table 4.3 LCI_{2.5} and LCI₅ T test results for CF and HC participants. Inn/SF₆ = Modified Innocor™ Sulphur hexafluoride device. ExD/N₂ = Exhalyzer D Nitrogen device. LCI_{2.5} = Lung Clearance Index calculated from washout to 1/40th of the tracer gas starting concentration. LCI₅ = Lung Clearance Index calculated from washout to 1/20th of the tracer gas starting concentration.

	CF			HC			Between groups (Ino & ExD)	
	Inn/SF ₆	ExD/N ₂	p value	Inn/SF ₆	ExD/N ₂	p value	p value	p value
LCI_{2.5}	9.51±2.27	14.01±3.72	p=0.0001	6.3±0.51	7.3±0.72	p<0.0001	<0.0001	<0.0001
LCI₅	6.81±1.4	7.92±1.7	p<0.0001	5.08±0.37	5.09±0.31	ns	<0.0001	<0.0001

Similar differences also existed for FRC (Figure 4.5) and CEV between devices and between HC and CF patients (Table 4.3) but the ExD/N₂ did not demonstrate a statistical difference in FRC between HC and CF.

Respiratory Rate (RR) and tidal volume (V_T) were higher in Inn/SF₆ compared to ExD/N₂ except in HC where V_T was not statistically different between devices (Table 4.3). I also counted RR and V_T every 30 seconds throughout a sample of 10 CF patient MBW traces to see if the longevity of the washout in N₂ would change breathing dynamics across the washout. There were no differences in RR or V_T between 30-second periods of a washout (results not shown).

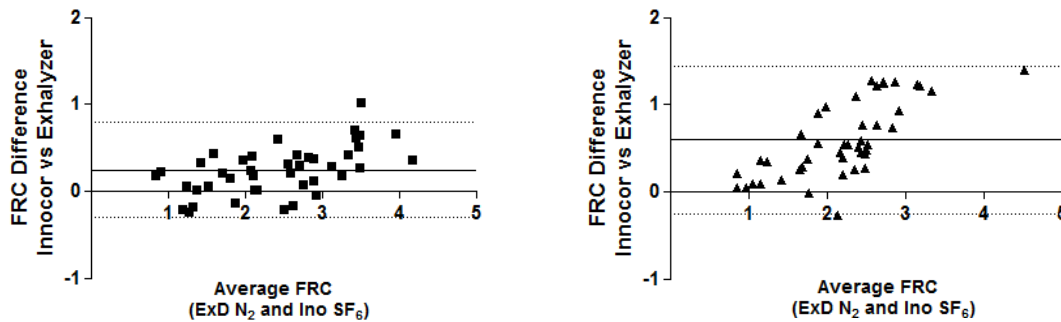


Figure 4.5 Bland Altman FRC differences between Inn/SF₆ and ExD/N₂ in HC on the left and CF on the right. ExD/N₂ FRC was higher than Inn/SF₆ FRC in both groups. The difference increased with higher FRC's.

Table 4.4 Modified Innocor and Exhalizer D T test differences in FRC, CEV, Respiratory Rate and tidal volume in both healthy control and cystic fibrosis patients.

	CF			HC			Between groups p values	
	Inn/SF ₆	ExD/N ₂	p value	Inn/SF ₆	ExD/N ₂	p value	Inn/SF ₆	ExD/N ₂
FRC (L) median (range)	1.91 (0.75-3.83)	2.53 (0.88 – 5.22)	<0.0001	2.44 (0.76-3.99)	2.6 (0.94 – 4.34)	<0.0001	0.01	ns
CEV (L) median (range)	38.43 (7.02 to 74.63)	18.47 (5.1 to 40.51)	<0.0001	19.8 ± 7.9	14.51 ± 5.6	<0.0001	0.009	<0.0001
Respiratory Rate (breaths per minute) Mean ± SD	18.95 ± 5.4	20.15 ± 6.2	0.02	14.4 ± 5.03	16.75 ± 5.6	<0.0001	0.009	0.0001
Tidal volume (ml) Mean ± SD	502 ± 154	540 ± 178	0.02	606.9 ± 233	594.6 ± 265	ns	ns	0.02

As shown previously LCI_{2.5} was statistically different between devices but in contrast, LCI₅ at an earlier time point was not (Table 4.3). This analysis shows that the discrepancy between washout traces is at the tail end of the washout.

CF LCI values obtained from Inn/SF₆ and ExD/N₂ were correlated with FEV₁ (r^2 0.45, $p < 0.0001$ and r^2 0.48, $p < 0.0001$ respectively, Figure 4.6). The slopes differed significantly (Figure 4.6; $p = 0.03$) being steeper with ExD/N₂ and with further divergence with severity. This difference in slopes was not observed for LCI₅ (Figure 4.6) confirming that the difference between devices is related to the latter part of the washout curves. For FEV₁ both the percentage predicted and Z scores are reported, for ease of understanding and elimination of bias of age, gender and height with Z scores (Figure 4.7).

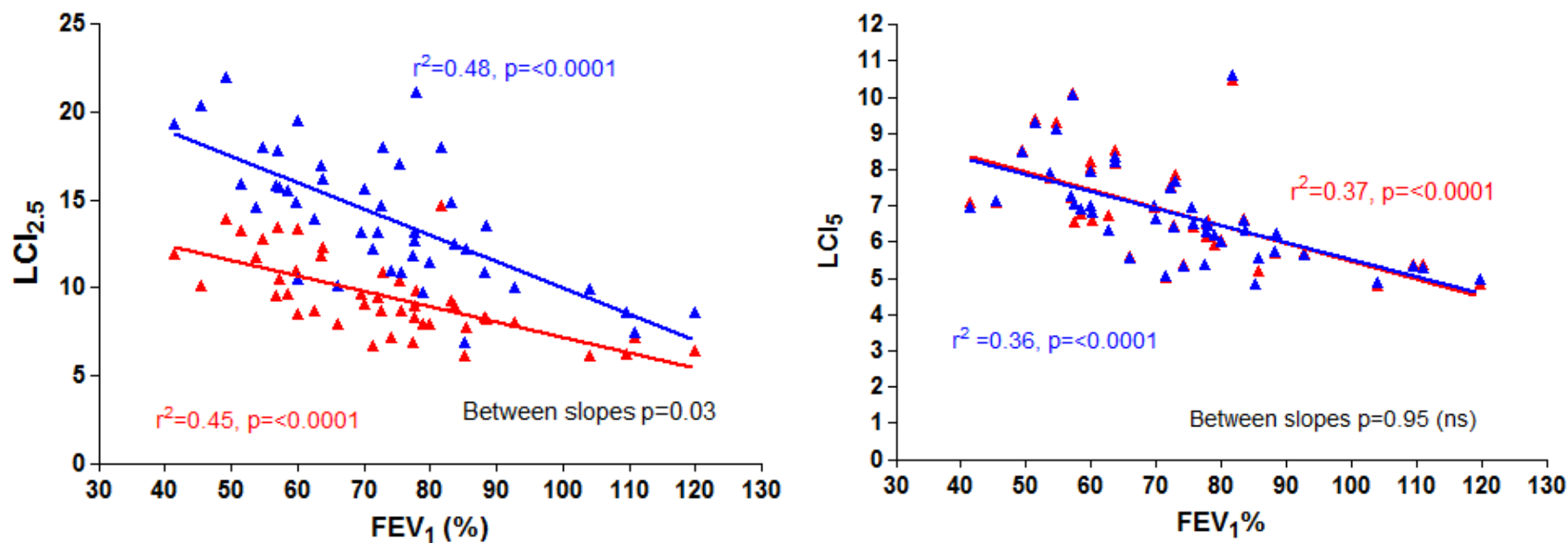


Figure 4.6 CF patients, plot of LCI_{2.5} vs. FEV₁% on the left and LCI₅ vs. FEV₁% on the right, from the Inn/SF₆ (red) and the ExD/N₂ (blue). Please note the difference in scales on all axes. All of the r^2 values for each device are statistically significant (LCI_{2.5} $r^2=0.45$ for Inn/SF₆, $p<0.0001$ and $r^2=0.48$ for ExD/N₂, $p<0.0001$ & LCI₅ $r^2=0.37$ for Inn/SF₆, $p<0.0001$ and $r^2=0.36$ for ExD/N₂, $p<0.0001$). In LCI_{2.5} the regression slopes are significantly different from each other ($p=0.03$) but they are not significantly different from each other in LCI₅ ($p=0.95$, ns).

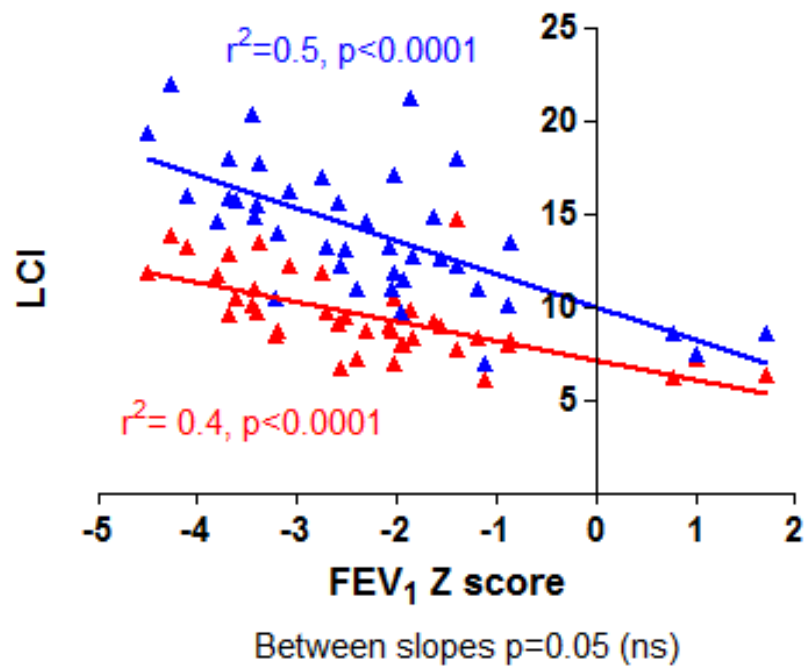


Figure 4.7 Plot of LCI_{2.5} vs. FEV₁ Z scores from the Inn/SF₆ (red) and the ExD/N₂ (blue). Both of the r^2 values for each device are statistically significant ($r^2=0.4$ for Inn/SF₆, $p<0.0001$ and $r^2=0.5$ for ExD/N₂, $p<0.0001$) and the difference between regression lines slopes is approaching statistical significance ($p=0.05$).

Illustration of the SF₆ and N₂ washout difference

The difference between washout slopes in one CF patient (Female, age 8, FEV₁% 88%) can be seen in Figure 4.8. In the SF₆ washout, LCI_{2.5} is calculated as 7.98 and the trace ends after just 12 breaths (A), whereas her N₂ LCI_{2.5} is markedly higher at 13.09 and the trace ends after 44 breaths (B). The difference between SF₆ and N₂ LCI₅ is less marked but LCI₅ is still higher with the N₂ washout (5.85 in SF₆ and 8.05 in N₂).

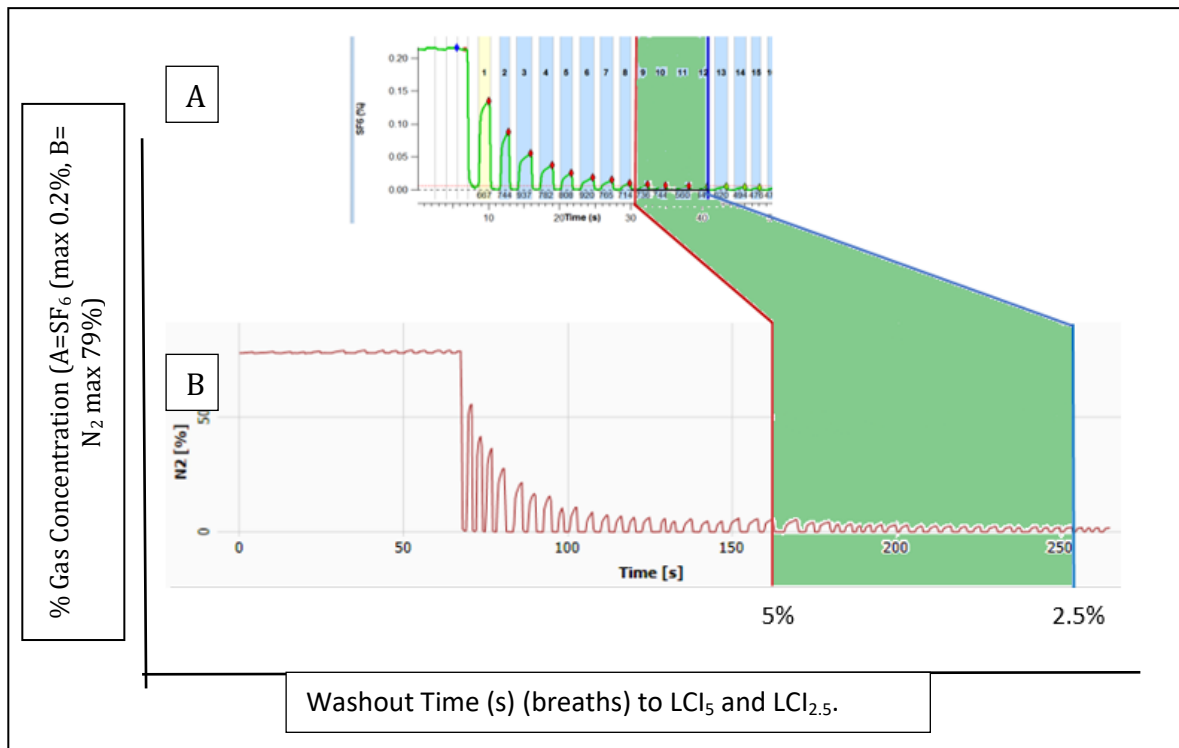


Figure 4.8 Individual illustration of the discrepancy seen between Inn/SF₆ LCI₅ to LCI_{2.5} (A) and ExD/N₂ LCI₅ to LCI_{2.5} (B) (CF, Female, Age 8, FEV₁% 88%). Calculated SF₆ LCI_{2.5} is 7.98 and ends after just 12 breaths, LCI₅ is 5.85 and there are 4 breaths remaining to get to LCI_{2.5}. On the same person, calculated N₂ LCI_{2.5} is 13.09 and it takes 44 breaths to achieve this end gas concentration. There are 23 breaths to get to LCI_{2.5} from an LCI₅ of 8.05. The difference between SF₆ and N₂ is shown in green.

4.4.2 Examine whether the differences between devices are impacted by disease severity (Aim 2)

Methods: LCI is largely considered to be more useful in milder disease than more severe disease where spirometry may provide a useful read-out [107]. FEV₁% and FEV₁ Z scores were compared with LCI results in groups of differing disease severity using linear regression analysis. 44 subjects with Inn/SF₆ and ExD/N₂ data were split into 3 groups, FEV₁% predicted <60% (n=11), 60-80% (n=21) and >80% (n=12). A one-way ANOVA was used to compare groups overall with a Bonferroni multiple comparisons test to see which groups were different.

In light of the devices showing different LCI results I hypothesised that the regression lines would also diverge, this would be particularly true in more severe disease.

43 extra CF patients with ExD/N₂ data from one time point (majority of the patients who took part in the treatment study in Chapter 6) were then added to the ExD/N₂ data to increase sample number and examine any differences between severity banding. To show the extent of disease that LCI could explain as well as understand differences between devices, linear regression analysis of FEV₁ and LCI was completed.

Results: The demographic data are shown in Table 4.5. 43 extra CF patients with ExD/N₂ data from one time point were then added into the data for a broader range of values to be available. Data from both devices were compared using paired analysis, and then all age groups were compared by linear regression.

Table 4.5 Cystic Fibrosis patient demographics for groups split by severity <60% (n=11), 60-80% (n=21) and >80% (n=12) FEV₁ predicted.

44 Patients with Innocor and Exhalyzer D data	<60%	60-80%	>80%
Number of subjects	11	21	12
Gender n (%)	10 (91%) Female	10 (48%) Female	5 (42%) Female
Age Mean $\pm$ SD years	27.5 $\pm$ 14.4	21.9 $\pm$ 9.5	15.4 $\pm$ 12
LCI ExD/N₂ Mean $\pm$ SD	17.3 $\pm$ 2.3	13.6 $\pm$ 2.8	11.3 $\pm$ 3.2
LCI Inn/SF₆ Mean $\pm$ SD	11.6 $\pm$ 1.5	8.9 $\pm$ 1.5	8.3 $\pm$ 2.3
FEV₁% Mean $\pm$ SD	52.5 $\pm$ 5.5	71.1 $\pm$ 3.8	92.7 $\pm$ 13.6
FEV₁ Z score Mean $\pm$ SD	-3.8 $\pm$ 0.4	-2.4 $\pm$ 0.4	-0.6 $\pm$ 1.1
43 CF patients added with Exhalyzer D data (84 patients total)			
Number of subjects	23	34	27
Gender n (%)	14 (61%) Female	16 (47%) Female	8 (30%) Female
Age Mean $\pm$ SD years	30.1 $\pm$ 15.3	22.2 $\pm$ 11.8	13.0 $\pm$ 7.5
LCI ExD/N₂ Mean $\pm$ SD	18.7 $\pm$ 3.5	13.4 $\pm$ 3.6	10.0 $\pm$ 2.2
FEV₁% Mean $\pm$ SD	51.4 $\pm$ 6.4	69.2 $\pm$ 0.5	89.3 $\pm$ 8.3
FEV₁ Z score Mean $\pm$ SD	-3.8 $\pm$ 0.5	-2.5 $\pm$ 0.1	-1 $\pm$ 0.5

There were LCI differences between all severity groups in both devices using an ANOVA test (ExD/N₂ p=0.0001 and Inn/SF₆ p=0.0002). Post Hoc analysis revealed the differences were between all severities with ExD/N₂; with Inn/SF₆, the significant ANOVA was generated by differences between <60% and 60-80% whereas 60-80% and >80% were not statistically different.

ExD/N₂ results were consistently higher than Inn/SF₆ in all severity groups (<60% p<0.0001, 60-80% p<0.0001 and >80% p=0.02) (Figure 4.9).

Interestingly when using FEV₁% on the x-axis there was a significant linear regression line for the ExD in <60% and >80% but not for 60-80%. In the Ino the opposite was true where 60-80% had a positive linear regression line but <60% and >80% were not statistically significant (Figure 4.9). However, when using Z scores on the x-axis, only the linear regression line for ExD/N₂ was significant in the >80% group (Figure 4.10).

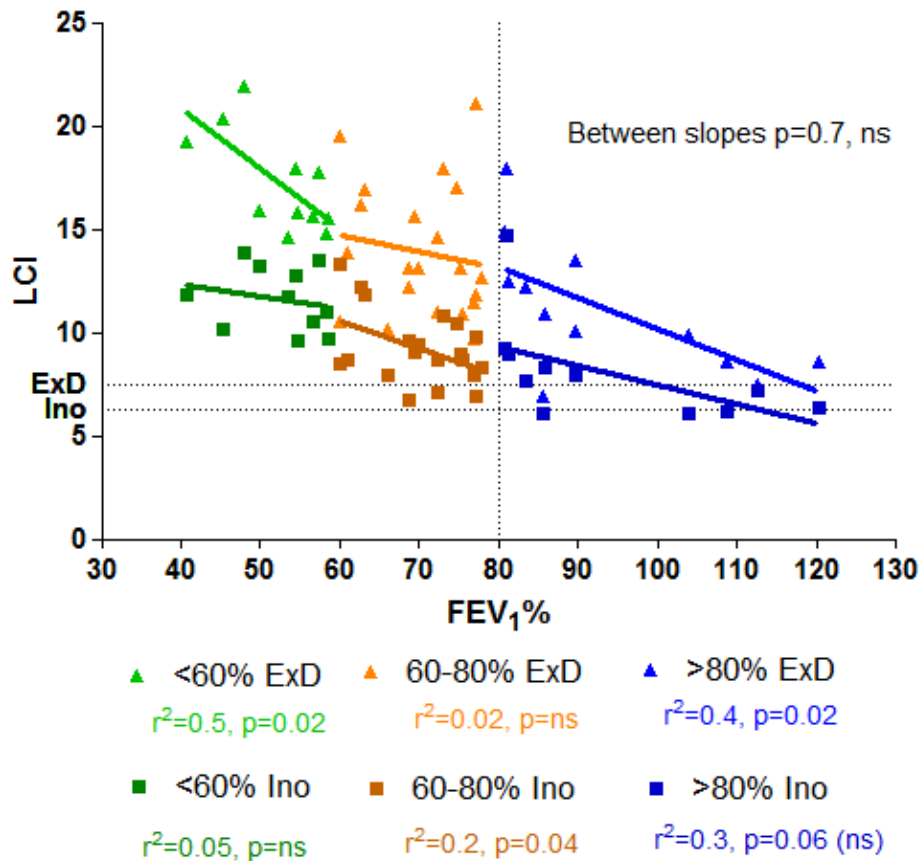


Figure 4.9 Graph to show the linear regression of FEV₁% and LCI from the Exhalyzer D (triangles) and modified Innocor (squares) devices in 44 CF patients, split into <60% (green), 60-80% (orange) and >80% (blue) groups. The linear regression lines between devices at each severity and a 3-way comparison across severities for each device are not statistically different from each other.

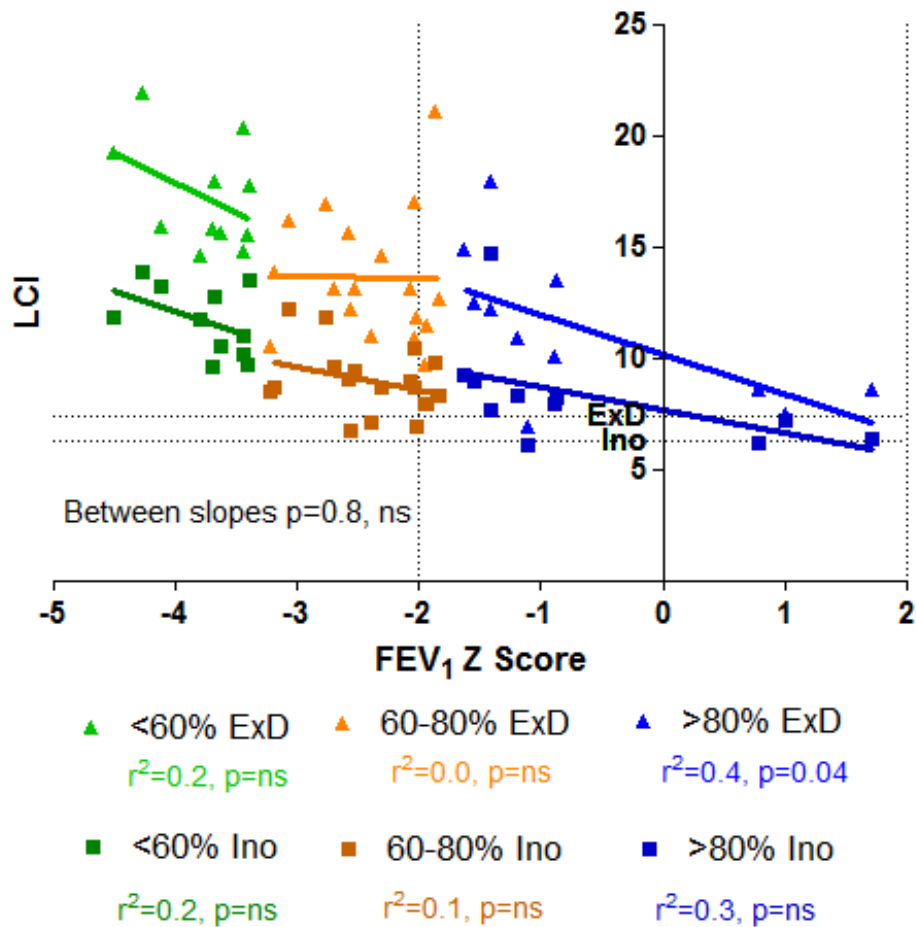


Figure 4.10 Graph of the linear regression between FEV₁ GLI Z scores and LCI (HC data from this thesis) from the Exhalyzer D (triangles) and modified Innocor device (squares). 44 CF patients, split into <60% (green), 60-80% (orange) and >80% (blue) groups. The linear regression lines for each group are all significant but the regression slopes are not different from each other when using Z scores ($p=0.8$, ns).

To increase the power of the analysis more CF patients with ExD/N₂ results had their FEV₁ (% and Z score) compared with LCI. By increasing the n number for ExD/N₂ there were significant regression lines in <60%, 60-80% but not >80% when using FEV₁% as the x axis, only the 60-80% group had significant regression line in FEV₁ Z scores (Figure 4.11). This highlights the differences between reporting physiological data in percentage predicted vs. Z score.

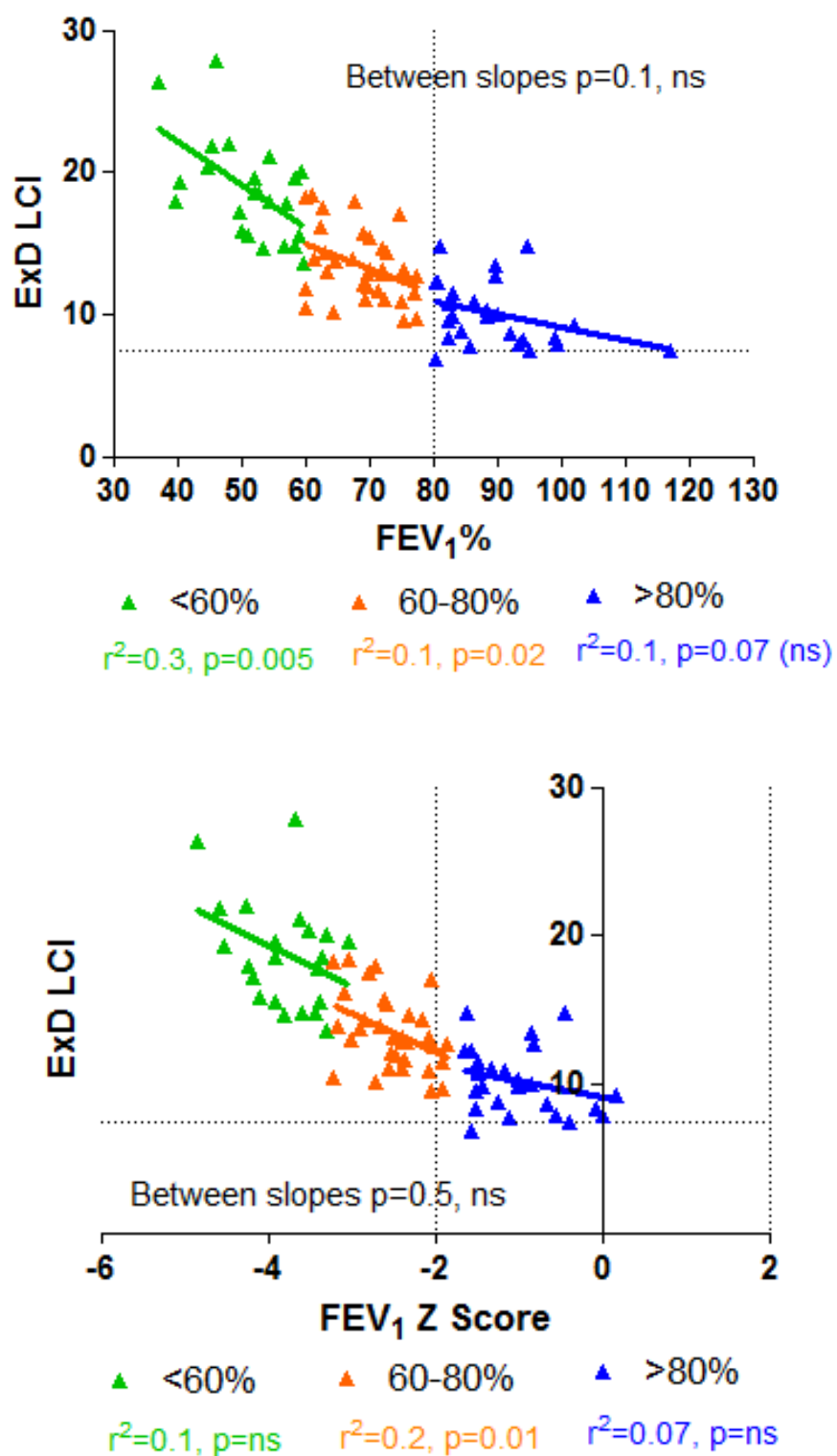


Figure 4.11 Graph to show the linear regression of FEV₁% and FEV₁ Z score against LCI from the Exhalyzer D device in 84 CF patients, split into <60% (green), 60-80% (orange) and >80% (blue) groups. The linear regression lines across severities for each device are not statistically different from each other. The linear regression lines were significant <60% and at 60-80% but not >80% (likely due to few patients with LCI within the normal range).

4.4.3 Understanding LCI differences between gases: (Aim 3)

4.4.3.1 Does FRC cause the difference between devices?

FRC derived from N_2 has previously been shown to be similar to that obtained by plethysmography [118], this is surprising since N_2 is derived from a washout method and therefore should be smaller than body plethysmography that encompasses all gas volumes [65]. The paper suggests that N_2 washout may be a more accurate reflection of FRC when compared to FRC calculated by SF_6 as it is similar to FRC_{PLETH} but other papers suggest it is likely to be discrepancies within the FRC calculation from using specific N_2 washout devices [165].

The aim of this section was to compare FRCs derived from SF_6 and N_2 washout with FRC calculated from recent body plethysmography. These answers will help to resolve the differences in FRC calculation from different devices. I hypothesize that FRC from N_2 will be higher than FRC from SF_6 but both will be lower than FRC from body plethysmography.

Methods: 11 HC and 24 CF patients from the overarching comparison study had undergone body plethysmography within a year of completing the N_2 and SF_6 MBW comparison. The FRC values from both devices were compared to FRC from body plethysmography.

The Ino and ExD MBW tests were completed on the same day in an alternate order. For the CF subjects, the body plethysmography used was the most recent result (within a year) as part of an annual assessment for CF. CF patients are required to be stable when attending their annual assessment visit (as they were for their washout visit) so this was considered a reasonable compromise as I had neither capacity nor ethical permission to undertake an additional test for research purposes. 15 of the cohort were children however and for convenience annual assessment data is sometimes collected at the end of an inpatient stay so results may reflect optimal health and this should be considered. For HC, staff within the Royal Brompton Hospital Lung Function department completes body plethysmography routinely for standardisation of equipment and to accrue biological control data. I completed MBW tests on these HC volunteers for FRC comparison.

A paired repeated measure ANOVA and post hoc Bonferroni testing assessed differences between devices.

Results: As described, LCI is calculated by dividing CEV by FRC, so if either of these parameters were different between devices then it could explain the LCI difference seen

above. This section compared FRC results from the Inn/SF₆ device, ExD/N₂ device and body plethysmography.

Table 4.6 Demographic data from HC and CF patients where FRC results from Inn/SF₆, ExD/N₂ and body plethysmography were collected and compared.

Table 4.5	Healthy Controls (11)	Cystic Fibrosis (24)
Gender n (%)	8 (73%) Female	15 (63%) Female
Age (years) median (range)	28.4 (20.4-56.1)	15.8 (7.7-63.7)
Height (cm) Mean $\pm$ SD	168.2 $\pm$ 9.5	154.2 $\pm$ 17.0
Weight (kg) Mean $\pm$ SD	65.6 $\pm$ 14.1	53.1 $\pm$ 20.0
FEV₁% (GLI [57]) Mean $\pm$ SD	Not Done	76.1% $\pm$ 17.18
FEV₁ Z Score (GLI [57]) median (range)	Not Done	-2.06 (-3.7-1.7)

In HC, FRC from the 3 devices were not statistically different from each other (FRC_{INO} 2.87 $\pm$ 0.48, FRC_{EXD} 3.09 $\pm$ 0.6 and FRC_{PLETH} 2.93 $\pm$ 0.49, p=ns – Figure 4.12).

In CF, FRC from the three devices were statistically different (FRC_{INO} 1.89 $\pm$ 0.67, FRC_{EXD} 2.41 $\pm$ 0.96 and FRC_{PLETH} 2.54 $\pm$ 0.85, ANOVA p=<0.0001 – Figure 4.12). Post hoc analyses revealed that FRC_{PLETH} and FRC_{EXD} were similar to each other (p>0.05) but FRC_{INO} differed significantly from those obtained with both of the other devices (p<0.001). These results confirm that FRC from SF₆ and N₂ are different from each other but the hypothesis that N₂ FRC will be smaller than FRC from plethysmography can be rejected.

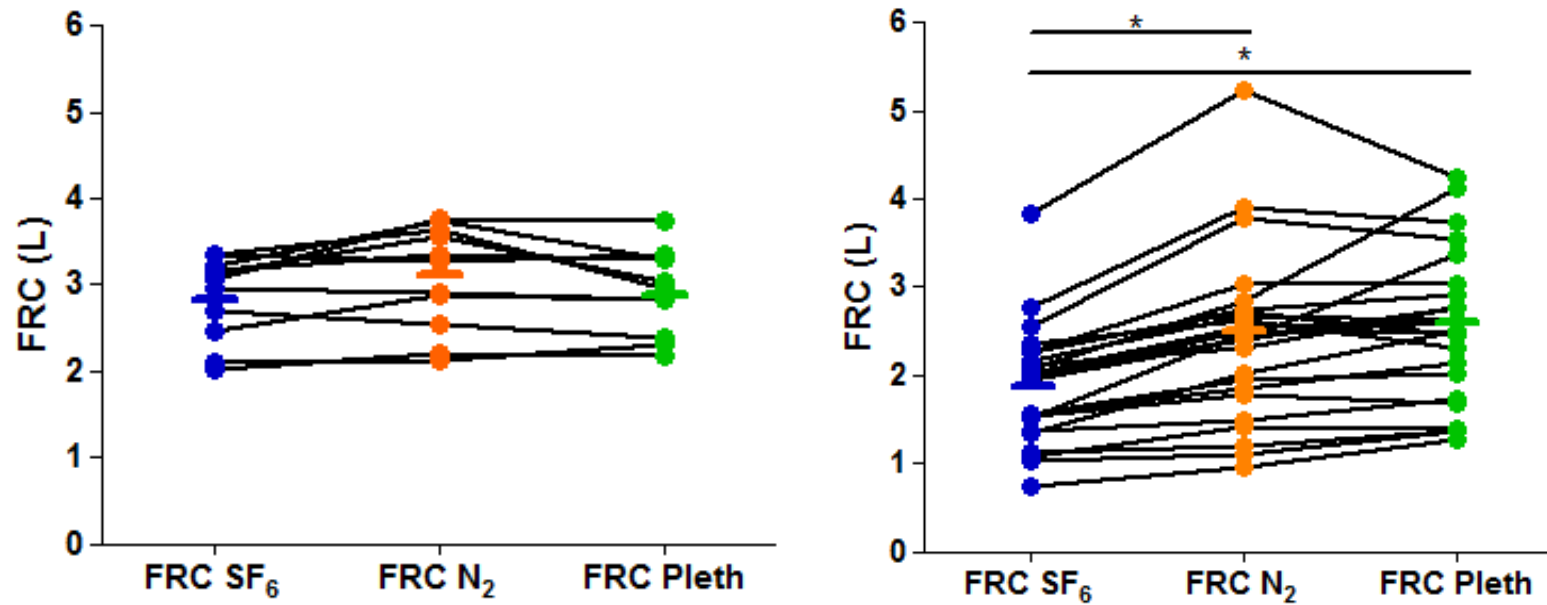


Figure 4.12 HC and CF FRC results obtained from the Innocor SF₆ device (blue), Exhalyzer D N₂ device (orange) and the FRC from body plethysmography (green). ANOVA Results were not significantly different between devices in HC. In CF FRC ANOVA results were different between devices, $p < 0.0001$, post hoc Bonferroni multiple comparisons analysis showed the difference was between SF₆ and N₂ ($p < 0.001$) and SF₆ and pleth ($p < 0.001$) but not between N₂ and pleth ($p = ns$).

4.4.3.2 Do individual patient factors between tests contribute to the difference? Simultaneous washout with an extended washout period

An additional explanation for the differences between the Inn/SF₆ and ExD/N₂ results could be technique-related differences in an individual's respiratory pattern or breath volume. This section will describe the design of a novel set-up to perform simultaneous washout; I hypothesised that, as both gases were washed out at the same time, I would be able to eliminate individual differences if it were not contributing.

In order to complete an SF₆ and N₂ washout at the same time, 100% oxygen was used as the washout gas (Figure 4.13).

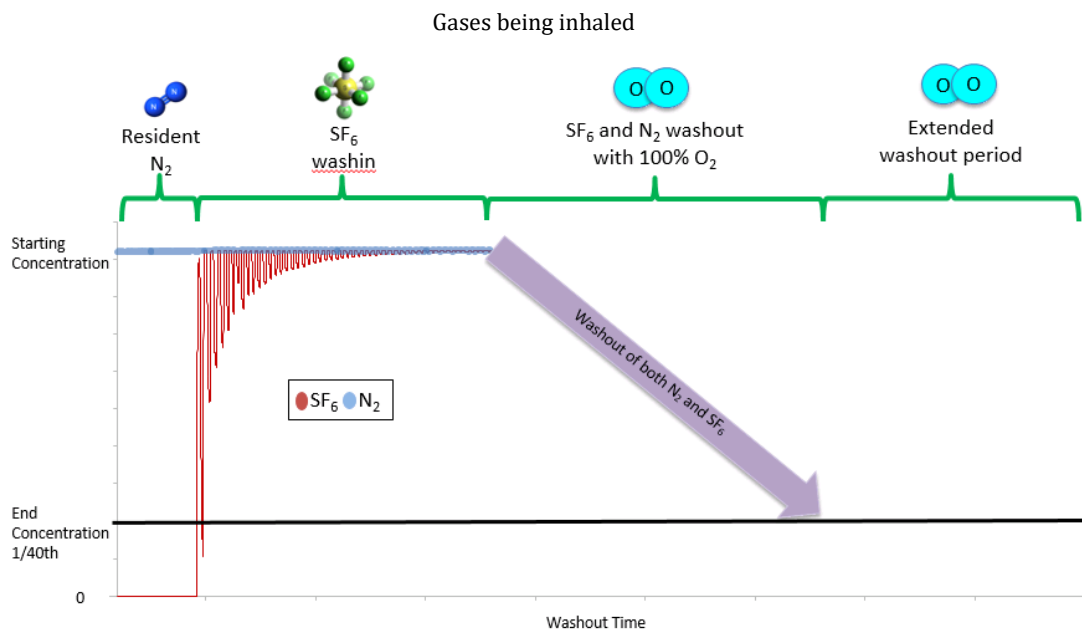


Figure 4.13 Diagrammatic representation of the washin of SF₆ (red) followed by simultaneous washout of SF₆ and N₂ (blue). The gases being inhaled during each washin or washout phase are highlighted along the top of the graph.

Methods: Healthy control subjects and a small sample of CF participants were recruited for this section.

Simultaneous washout of SF₆ and N₂ using was achieved by attaching the ExD/N₂ and Inn/SF₆ devices in series (see Figure 2.9 in the methods section). The participant completed tidal breathing whilst washing in 0.2% SF₆ gas cylinder. When equilibrium of 0.2% SF₆ in the lungs was detected by the Innocor device (normally around 2 minutes in healthy individuals) the “Start WO” button on the

Exhalyzer D was pressed. Gas tubing was manually swapped from the 0.2% SF₆ to the Exhalyzer D tubing. 100% O₂ (electronically switched on by the Exhalyzer D) was then breathed for the washout of 0.2% SF₆ and ~79% N₂.

Washout of both gases was detected down to 2.5% of the starting concentration (SF₆ monitored on the Innocor device and N₂ on the Exhalyzer D). The participant then continued to breathe 100% O₂ for a further 2 minutes. The extended washout period was to observe how low a gas concentration could be achieved beyond the 1/40th end tidal gas concentration and the shape of the gas decay curve. Raw gas concentration, breath number and time to washout were recorded and analysis was performed (T tests or Wilcoxon signed ranked tests) after exporting the raw breath tables from both analysis softwares and matching each breath with the decay of N₂ and SF₆.

Results: Demographic data is displayed in Table 4.7.

Table 4.7 Demographic data of subjects completing a simultaneous washout of SF₆ and N₂.

(Inn/SF ₆ & ExD/N ₂ in series)	CF (n=5)	HC (n=8)
Gender number (%)	4 (80%) Female	5 (63%) Female
Age (years) (median (range))	24 (12-51)	31.25 (23-50)
Height (cm) Mean ± SD	160.4 ± 11.13	171 ± 7.66
Weight (kg) Mean ± SD	60.5 ± 18.23	74.2 ± 20.1
FEV₁% Mean ± SD, Z Score	84.02 ± 13.3, -1.3 ± 0.93	Not Done
Mean ± SD		

When simultaneous washout of 0.2% SF₆ and N₂ using 100% O₂ was performed with the Inn/SF₆ and ExD/N₂ attached in series, clearance of test gas to 1/40th still occurred after fewer breaths for SF₆ than N₂ in both CF (CF 25.5 (14-35.5) vs. 54.5 (38-107.5) breaths, p=0.06, ns) and HC (46.3±13.7 vs. 32.04±11.94 breaths p=0.002). The lack of significance in CF is likely due to lack of power. The fold impact was significantly greater in the CF group (2.5(0.9) fold), than the HC (1.5(0.3) fold) (p<0.05). Thus, the difference between washouts cannot be accounted for by any changes in respiratory rate or tidal volumes, which have been corrected for here.

Examples of the decline in tracer gases for HC and CF are shown in Figure 4.14, the gas concentrations have been plotted on a percentage normalised log scale so they can be seen together. SF₆ and N₂ concentrations decline similarly at the beginning of the washout but separate beyond approximately 20 breaths and continue to diverge at the tail end. SF₆ continues linearly downwards whereas N₂ curves, eventually reaching a plateau in all subjects (mean concentrations

1.26(0.09)% in HC and 1.48(0.12)% in CF) from which it does not decline further, this could represent a tissue N_2 contribution in the N_2 washout.

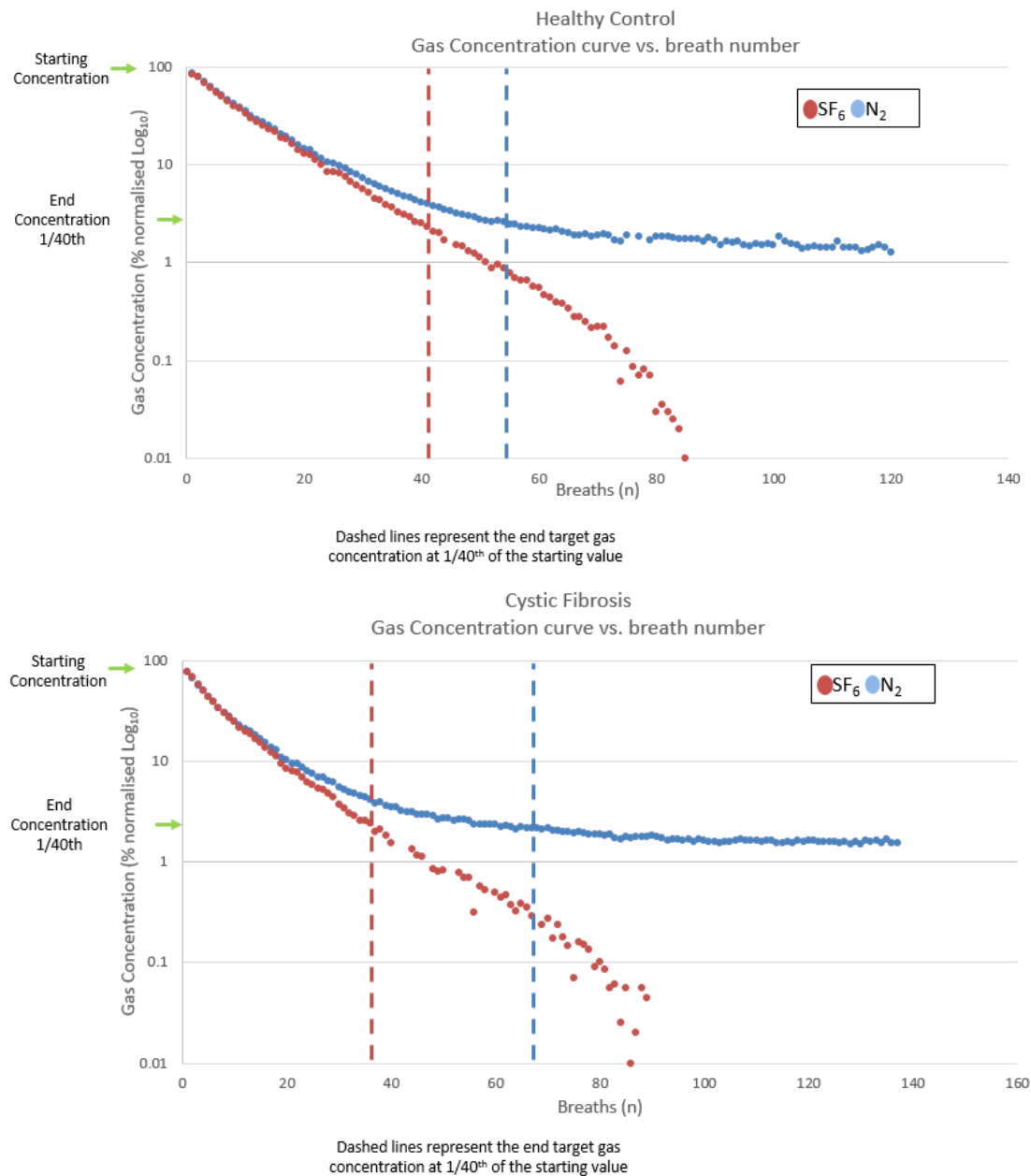


Figure 4.14 Graphs of a healthy control and cystic fibrosis individual's gas concentration decline throughout a simultaneous washout of sulphur hexafluoride (SF_6 , red) and Nitrogen (N_2 , blue) vs. breath number. The gases are displayed in a normalised log scale so that each gas concentration is shown on the same graph. The vertical lines (dashed red = SF_6 , dashed blue = N_2) represent the end tidal gas concentration at 1/40th of the starting concentration, beyond this point shows the extended washout period. In all subjects, SF_6 reached the end target first. N_2 seemed to reach an asymptote around 1% N_2 whereas SF_6 continued down to zero.

4.4.3.3 Respiratory Mass Spectrometry

This section is split into two, the first exploring the impact of breathing 100% oxygen versus room air to washout N_2 or SF_6 and the second questioning whether differences in N_2 and SF_6 washouts can be explained by differences between equipment.

4.4.3.3.1 Do tracer gases washout differently? Does breathing 100% O_2 impact the washout?

One question that was raised after completing the simultaneous washout is whether or not the washout of SF_6 is affected by using room air or 100% O_2 ? The N_2 washout could also have been impacted by the addition of SF_6 .

Is the washout of SF_6 different when using 100% O_2 compared to room air?

Is the N_2 washout affected by the addition of SF_6 ?

The methods section describes the RMS and specific methodology utilised in this thesis. The RMS gas sample line measured the gases from a washout whilst the operator measured time and breath number to washout completion. The different deadspace and sample collection from each device can be seen in Figure 2.14 in methods.

7 HC participants completed four SF_6 MBW tests using the RMS (SF_6 4%). Two tests used room air as a washout gas and two used 100% O_2 as a washout gas (allowing monitoring of both SF_6 and N_2 washout).

Those same 7 HC's also completed a conventional N_2 washout (100% O_2) and results were compared to the simultaneous N_2 and SF_6 washout with 100% O_2 . Results were compared using student t tests in Graphpad Prism software.

Results: During simultaneous washout of SF_6 and N_2 , 100% O_2 is utilised; these results can therefore show whether an SF_6 washout is different when using room air or 100% O_2 . The results also show whether an N_2 washout is impacted by the addition of SF_6 .

1. When washing out 4% SF_6 , on the RMS there were no differences in breath number when using room air or 100% O_2 (Figure 4.15).
2. When washing out N_2 alone or N_2 with the addition of 4% SF_6 there were no differences in N_2 washout breath number (Figure 4.15).

This shows that simultaneous washout using 100% O_2 is valid for both gases.

Washout time was slightly higher when both gases were washed out together for N_2 but there was no difference in SF_6 testing. RR was similar between tests in all scenarios (data not shown).

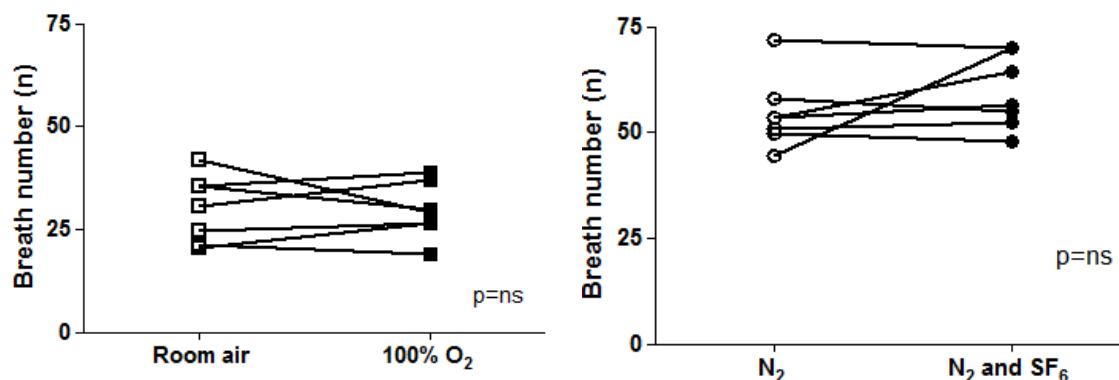


Figure 4.15 Washout breath number for 7 HC subjects completing 4% SF₆ washouts using room air or 100% O₂ on the left and N₂ washout with 100% O₂ or N₂ and SF₆ washout with 100% O₂ on the right. T test analysis showed no significant differences between washout gases or washout breath number (Room air 29.98 ± 8.2 vs. 100% O₂ 29.57 ± 6.8 and N₂ 54.67 ± 8.7 vs. N₂ with SF₆ 59.5 ± 8.7).

4.4.3.3.2 Does using the respiratory mass spectrometer to measure both gases eliminate equipment differences?

The RMS allows direct measurement of all gas concentrations therefore using the RMS should resolve the question of whether SF₆ and N₂ washouts are inherently different from each other. By incorporating the Inn/SF₆ and ExD/N₂ devices then disparities linked to the differences in equipment might be able to be explained, this also negated the difficulties with RMS set up and not being able to record flow with the RMS alone. Appendix 2 shows methodological steps to the final RMS testing.

The ExD/N₂ and Inn/SF₆ were attached in series and the sample line from the RMS fixed in place in the breathing circuit (Figure 2.9 in the methods section). A specialist 1% SF₆ gas mix with 21% O₂ and balance N₂ was utilised so that all three devices could measure O₂ and CO₂; N₂ was measured by both RMS and ExD/N₂ and SF₆ by RMS and Inn/SF₆. Sensitivity to SF₆ on the RMS was increased and ATPC was turned on. Raw results for initial and final gas concentrations were recorded. Breath number and time to washout to 1/40th for the RMS was calculated manually by averaging the end tidal washout concentration for both SF₆ and N₂ breaths using the modified Innocor Simple Washout (SW) software. As flow was not collected on the RMS, calculation of FRC and LCI were not possible. Raw results were compared using student t tests.

A sample of results was taken to try to indirectly calculate the N₂ signal from the Innocor results. Problems including viscosity corrections, adjusting the Oxigraph to account for the addition of other gases and other issues discussed previously i.e. signal synchronisation and delay time; meant that this testing was discarded.

Results: Having demonstrated that even with simultaneous washouts, SF₆ required fewer breaths to reach the defined washout concentration of 1/40th than N₂, I aimed to remove any effect of equipment, gas analysers and software by completing a simultaneous washout of both SF₆ and N₂ on the RMS; both of the other two devices were connected to allow a direct comparison.

Table 4.8 Demographic data of HC subjects who completed the 1% SF₆ and N₂ washout. Inn/SF₆ = Modified InnocorTM Sulphur hexafluoride device. ExD/N₂ = Exhalyzer D Nitrogen device. RMS = Respiratory Mass Spectrometer.

(Inn/SF ₆ & ExD/N ₂ in series with RMS sample line)	HC (10)
Sex n (%)	6 Female (60%)
Age (years) (median (range))	28.27(21.4-35.3)
Height (cm) Mean ± SD	169.9 ± 10.4
Weight (kg) Mean ± SD	66.7 ± 9.9

Using RMS only, N₂ is slower to washout than SF₆ (33.07 ± 13.2 vs. 26.64 ± 11.01, p=0.02); it is likely that the gases are inherently different.

1% SF₆ washout testing with 100% O₂ increased the signal to noise ratio of SF₆ particularly towards the end of the washout (Figure 2.11 in the methodology section) as well as increasing the measurement error of detected gas concentrations. The starting concentration of each gas (1% SF₆ (+/-5% (0.95 to 1.05) and 78.02% N₂ (+/- 5% (74.12 to 81.92)) showed large deviations between devices; SF₆ (-16.25% in Inn/SF₆ and -11.36% in RMS) and N₂ deviation (ExD/N₂ -0.42% and RMS 1.52%). Therefore these results should be interpreted with care.

N₂ and SF₆ washout breaths from the ExD and Ino were statistically significant from the RMS, lower for RMS (SF₆ = Inn/SF₆ breaths = 31.21 ± 12.7 vs. RMS breaths 26.64 ± 11.01, p=0.001 and N₂ = ExD/N₂ breaths = 37.9 ± 15.95 vs. RMS breaths 33.1 ± 13.24, p=0.02) (Figure 4.16).

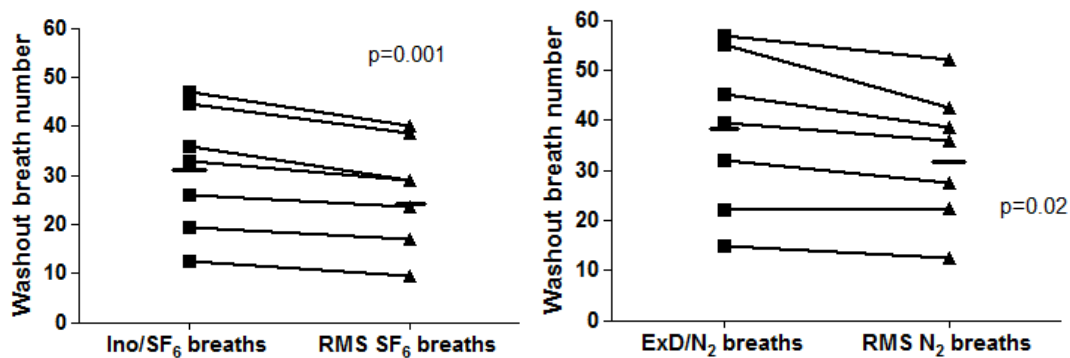


Figure 4.16 10 HC participants completed a simultaneous washout of 1% SF₆ and N₂ using the Innocor (Inn/SF₆) and Exhalyzer D (ExD/N₂) attached in series and the respiratory mass spectrometer (RMS) gas sample line placed in between. This plot shows the t test difference in washout breaths for sulphur hexafluoride (SF₆) (left) and nitrogen (N₂) (right) in Inn/SF₆ and ExD/N₂ vs. RMS. Breath number was always lower in the RMS compared to the other devices.

An individual example of SF₆ and N₂ washout decay from all of the devices is shown in Figure 4.17. The gases decay similarly up to breath 20 then begin to separate. Like in Figure 4.14, SF₆ washout completes before N₂ and there is a separation of SF₆ and N₂ towards the end of the washout. The extent of the difference between SF₆ and N₂ on the RMS (5.5 (3-12) breaths) vs. the delta difference between Ino and ExD (6.5 (2.5-10.5) breaths) was not significantly different. This is likely due to both the Inn/SF₆ and ExD/N₂ having been validated using the RMS.

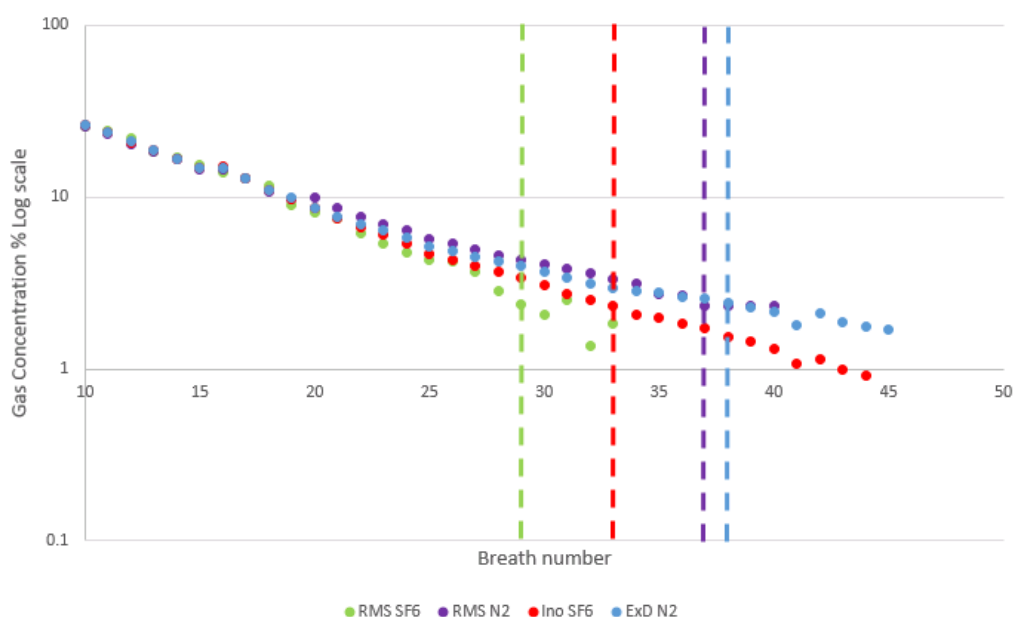


Figure 4.17 Graph of a healthy control simultaneous washout of SF₆ and N₂ on a normalised log scaled graph for the RMS, Innocor and Exhalyzer D. RMS SF₆ (green), Inn/SF₆ (red), RMS N₂ (purple) and ExD/N₂ (blue). The dashed lines represent the breath number where the washout reaches the 1/40th target. Gases decay similarly in the first instance and separate beyond breath 20 (breath 1-10 not shown to show separation of the gases).

4.5 Discussion

This chapter focussed on the differences between N₂ and SF₆ after completing a comparison study and finding that N₂ LCI results were higher than LCI results from SF₆. In summary, I found that N₂ washout is inherently different to SF₆, as the differences still persist when simultaneous washout occurs and when gases are analysed on just one device. The main factor is likely to be the contribution of tissue N₂ to the N₂ washout. The sections below discuss these findings in detail, highlighting strengths and weaknesses of the approaches I used and identifying areas requiring further work.

4.5.1 Comparison of N₂ and SF₆ LCI results

A comparison of an SF₆ washout on the modified Innocor device against a N₂ washout on the Exhalyzer D device was performed in the same individuals. With N₂ washout, LCI and FRC were higher than with SF₆ and the discrepancy was larger in CF, increasing with more advanced disease (Figure 4.4). The discrepancy between devices seen here was much larger than seen previously [118, 165]. The hypothesis that LCI results from both devices will correlate but differ in absolute value has therefore been satisfied. A meaningful clinically important difference (MCID) for LCI has not been officially established but treatment responses of -0.8 and -1.16 LCI units have been reported [123,

159]. These responses are smaller than the LCI unit differences demonstrated between these techniques (0.87 in HC and 4.2 in CF) underscoring the importance of these findings. N₂ results were also higher than SF₆ in a similar comparison study reported previously [118] but importantly have been completed on different devices showing that the discrepancy between SF₆ and N₂ washout gas is much larger here and therefore is not simply related to device.

The MBW consensus statement highlights that tests with LCI over 10% or FRC with over 25% variability should have their acceptability reviewed [101]. The results here show that both LCI and FRC CV% in Inn/SF₆ and ExD/N₂ are within acceptable limits (Table 5.3) and, reassuringly, are not different to each other. Steps to reduce individual differences were taken i.e. alternating the order of testing and promoting similar testing experiences and the same appropriately calming DVD being played throughout testing so individual differences should be minimised. However, as the difference between devices was so large, patient factors including test sequencing, change in breathing pattern and device preference cannot be ignored.

The comparison showed that other parameters of the MBW washout were also significantly different between devices. Some have suggested that breathing 100% O₂ affects the breathing dynamics of individuals [213] and this may in turn impact the difference between MBW techniques. FRC is calculated through tidal breathing, if breathing pattern over time is variable then FRC will not be true and subsequent LCI calculation will not be an accurate reflection of ventilation heterogeneity [216]. V_T in this study was not different in HC but there was a small significant difference in V_T and RR between devices in CF. In a sample of patients RR and V_T were assessed in 30 second periods throughout the washout but there were no differences over time, this implies that the duration of the washout and potential more relaxed breathing over time was not a cause for the difference and 100% O₂ may indeed influence breathing dynamics. Without looking at a simultaneous washout of SF₆ and N₂ these changes to breathing pattern cannot be entirely disregarded (please see section 4.5.3.2).

This section of the thesis highlights the differences between N₂ and SF₆ washouts but it does not provide evidence of the underlying mechanisms for the differences. Reassuringly LCI from the Inn/SF₆ and the ExD/N₂ are able to distinguish between HC and CF patients suggesting that LCI is a robust tool despite differences between devices. N₂ and SF₆ do not give the same LCI results; N₂ results are higher. MBW results cannot be used interchangeably and I believe that LCI results should be reported with the equipment and test gas used so that discrepancies can be identified and misinterpretation of results are avoided.

4.5.1.1 LCI_{2.5} and LCI₅

The difference between gases occurs at the tail end of the washout where the small and poorly ventilated lung units contribute to the washout slowly. This can be shown by the separation of gases in LCI_{2.5} that is not seen in LCI₅ (Figure 4.6) (this can be seen by the area highlighted in green in illustration 1 (Figure 4.8)). LCI_{2.5} incorporates a larger proportion of inhomogeneity from the lung as the washout gas is detected to a lower concentration. LCI₅ has been shown to decrease test time and potentially make MBW easier for participants in a clinical setting [166, 185] however I show here that LCI₅ does not account for all of the inhomogeneity within the lung and this may be important disease information that is missed. Time saved, particularly in young children, is not long, with only a few more breaths required to reach the full 1/40th washout and Hannon *et al.* 2014 [167] suggest that the clinical utility of LCI, because of the time consuming nature of MBW in adults, is limited anyway. Full MBW testing to 1/40th of the starting concentration should be performed, this way all of the information available from a washout can be revealed and future expansion on other MBW variables such as phase III analysis can be explored to their full potential. The potential sources contributing to the differences between SF₆ and N₂ are offered below.

Poncin *et al.* 2016 suggest that the improved agreement at an earlier time point may be because the algorithm determining nitrogen fractions (washout tracer) could be subject to drift of a signal over time. A sample of patients here had ExD/N₂ results calculated at the 1/40th Innocor target (Data not shown). These results still showed discrepancies, so despite matching for breath; FRC and CEV were still higher in ExD/N₂ at this time point. This would suggest that it is not just N₂ software algorithms that contribute to the difference between devices but also the algorithms for CEV and FRC. It also shows the difference persists throughout the whole washout, not just the end.

4.5.2 LCI results and disease severity

As the discrepancy between devices increased in magnitude with a higher/more severe LCI, this section sought to explain differences between devices in terms of severity and attempt to define which cohort of patients would benefit the most from LCI measurement.

Overall

There were significant differences in the linear regression analysis between severity groups as well as between devices (Figure 4.9). The N₂ device was higher and unsurprisingly the lower the FEV₁% the higher the LCI, this implies that as disease worsens airway obstruction and ventilation heterogeneity is larger. I had expected that the mild cohort would show the most benefit of LCI use i.e. the shallowest regression line when comparing FEV₁ and LCI as LCI has been shown to be

abnormal in spite of normal FEV₁. However the linear regression line was not significant likely due to small patient sample when splitting by severity as well as the >80% cohort here had a mean age of 15.4 years rather than younger and disease onset will already have occurred. In line with the literature we know that issues with adherence, transition to becoming more self-reliant and increased exposure to clinical risk factors can all affect CF disease and variability and lung function decline at this age [18, 217].

Comprehensive studies have shown that LCI changes over time and this is not simply disease deterioration but the ageing of the lung as seen in healthy control individuals [143]. Lum *et al.* (2013) found that body size correction alone (FRC incorporated into the LCI calculation) was not sufficient to track LCI change in early childhood and could lead to clinically significant misinterpretation of results [120]. For these reasons it is unsurprising that results shown here are different between groups as well as the added disease severity. Comprehensive reference ranges are essential for more accurate interpretation of results (see methodology development section).

Does LCI provide additional information to FEV₁?

Figure 4.9 shows FEV₁% vs. LCI from three different severity groups. Not all of the regression lines are significant indicating that FEV₁% and LCI are not completely correlated and therefore one outcome measure may provide more disease information than another. Small patient numbers when splitting the cohort by severity is a limitation of this study and may have influenced this analysis.

The ExD/N₂ had significant regression lines in the most severe and most mild cohorts, this suggests that LCI and FEV₁% are in line with each other and there may be no additional benefit or identification of lung function abnormality when using LCI within this cohort. This is surprising since other literature shows that LCI is abnormal when FEV₁% is within normal limits. However LCI is abnormal in all patients where FEV₁% is above 80% so the benefit of LCI may be apparent but the small numbers and regression analysis may not identify this here. When looking at the 84 people with ExD data (Figure 4.11) 32% of patients could benefit from using LCI as an outcome measure rather than spirometry, these people all have mildly-impaired FEV₁. The correlation of two measures may not be the only measure of usefulness of a test however and though a test may correlate it may also provide complementary information to a result.

Is LCI useful once FEV₁ is abnormal?

The usefulness of LCI as a marker of disease progression once it is above the upper limit of normal is yet to be determined. In part this may depend on the degree of change to treatment and whether

the small airways disease that is detected by LCI is reversible. The longitudinal and individual variability of LCI also needs to be established further to understand if a single reading from one time point can be used to target treatment.

The disagreement between devices is increased disproportionately with disease, particularly above 10 LCI units (seen here and in Jensen *et al.* 2013). With increased numbers more investigation on the deflection of the slopes and correlation would be required as here the slopes were not significantly different from each other.

The minimum LCI value in the group with abnormal lung function is 8.3. Interestingly there are people in the top right hand quarter (Mild FEV₁ but abnormal LCI) with LCI's at this value or above (Figure 4.9). This shows the variability of disease between patients and perhaps the degree of variability within LCI. A larger number of patients are required to see if there is an LCI value, above which determines LCI is no longer useful. Until then FEV₁% and LCI could be used synergistically in these patients to track and monitor disease and prevent further progression.

Do the two devices give us the same message?

Interestingly there was no difference between the regression lines between devices in each severity group. This implies that the difference between devices is consistent when separating by severity therefore one device may not be superior to another in terms of read out capacity. For example, although the results are higher in ExD, for the same disease worsening there would be a similar increase in LCI on both devices. Due to the small numbers in each severity subgroup the divergence in worse disease seen in the whole group is not seen here, more testing, particularly at the mild and severe flanks would reveal further difference.

The spread of data in ExD/N₂ compared to the Inn/SF₆, may translate into improved power for clinical trials since there is more scope for change. Prevalence of abnormal LCI was greater in ExD/N₂ (>7.4 LCI ULN for ExD) compared to Inn/SF₆ (>6.3 LCI ULN for Ino) (41 out of 44 (93%) vs. 36 out of 44 (82%)), so the ExD/N₂ may be more useful for early identification of those patients at risk. This is a relatively small sample size however and patient variability could account for these differences. Potential overestimations in FRC and CEV in the ExD may also translate into the ExD/N₂ overestimating those with abnormal LCI compared to Inn/SF₆.

Unlike previous literature [1, 87] this dataset does not contain as many patients who have mild disease reflected in both LCI and Spirometry. This is probably due to the selected nature of the population in this thesis i.e. patients predominantly had a LCI performed in relation to starting a new treatment and therefore disease may already be beyond where LCI is normal. Both devices,

therefore LCI, identified abnormality in more patients than spirometry alone (32 out of 44 (73%), Figure 4.9) so use of either MBW device may be more useful than spirometry.

Z Scores

Z Scores are considered to be more robust than percentage predicted results because they encompass the normal distribution of a population [180]. It is interesting that there were fewer significant LCI regression lines when using FEV₁ Z score (Figure 4.10) compared to FEV₁%. This could be a factor of percentage scores not being as accurate as Z scores and therefore overestimating disease and producing a significant regression line compared to FEV₁ Z score. However the percentage results and Z scores are both produced with the GLI reference equations so they should be similar. With more patient numbers it is likely that the differences between percentage predicted and Z scores would be reduced. Ideally a comparison of FEV₁ Z score with LCI Z score would be completed but for reasons outlined in the methodology development section this was not done. The fact that only FEV₁ Z score and LCI in those mild patients using the ExD was significant (Figure 4.10) would suggest this device may provide a larger signal of inhomogeneity than Inn/SF₆. However this result is contrary to the hypothesis that LCI offers an advantage over FEV₁% as the regression lines are significant. It is also against the hypothesis that the regression lines will diverge in more severe disease as generally the regression lines between devices were similar, this could be do with splitting the group by severity therefore decreasing the sample number and decreasing the power; for this reason, any such conclusions need to be made cautiously and results confirmed in future, larger studies.

To increase the power of the study, further ExD/N₂ results were added to the linear regression analysis (Figure 4.11). In line with the hypothesis that the mild and severe group will not have regression lines that are significant, the only group to have a significant linear regression line was 60-80%. This shows that LCI offers different information to FEV₁ Z score; this is in line with the literature, where LCI has been shown to be abnormal despite normal FEV₁ [100]. The result could even show that LCI is also useful in these patients where FEV₁ is abnormal. Only further monitoring of LCI would reveal if LCI were useful in increased severity.

This analysis highlights the differences when using percentage predicted vs. Z scores and this should be considered when reporting data. Z scores are likely to be more robust as they incorporate covariates that have been shown to impact lung function [180] (see method development section). Interestingly when using Z score data from this thesis the two devices compare similarly to FEV₁ Z score, unlike when using percentage predicted. This highlights the requirement for large-scale standardised Z scores for LCI that could mean MBW results between devices would be comparable.

In conclusion, paediatric patients with milder disease are likely to benefit the most from the use of LCI as a marker of disease, however the sample of patients here may not show the advantages of LCI in mild disease as well since all patients had abnormal LCI results. LCI may also be useful in severe disease where differences between FEV₁ and LCI results are apparent. The differences between significance in regression lines are not clear-cut, particularly when using percentage predicted or Z scores but there were no differences in the correlation of FEV₁ and LCI between devices in any severity group; therefore both devices are equally useful for disease classification.

4.5.3 Understanding LCI differences between gases

4.5.3.1 FRC

With increased airway obstruction and gas trapping in CF, FRC from body plethysmography is likely to be increased [218]. In contrast to this, FRC from gas dilution methods (MBW) is likely to be diminished in CF patients, where increased gas trapping and airway obstruction mean that a smaller proportion of accessible/participating lung units are free to take part in gas exchange during the washout time [175]. This section shows that in HC, FRC measurements from both MBW techniques are similar to those obtained by plethysmography, likely due to little gas trapping amongst the healthy participants. However, FRC from ExD/N₂ is larger than FRC from Inn/SF₆ and body plethysmography in CF (Figure 4.12). These results are similar to those seen previously [118]. Two possible explanations have been considered:

- a) FRC_{N₂} provides a more accurate reflection of FRC as N₂ more easily accesses poorly ventilated areas of lung than SF₆ as it fails to washin.
- b) FRC_{N₂} is over-estimated in CF for reasons similar to those leading to longer washouts.

Jensen *et al.* 2013 suggests that the FRC from N₂ may provide an accurate reflection of ventilation heterogeneity similar to that obtained from body plethysmography. Body plethysmography is the gold standard of FRC measurement and incorporates the whole lung volume including gas trapping. When calculating FRC from the gas dilution technique, volume of trapped gas may be missed due to partial or inaccessible lung units that do not communicate within the washout time [65]. N₂ is a smaller gas molecule than SF₆ and one that is resident within the lung, therefore it may be that N₂ can reveal more ventilation heterogeneity and gas trapping (that is inaccessible to SF₆ only). The ability to measure and detect this mechanism is limited without imaging of gas trapping and mucus plugs; therefore, overestimation errors in the N₂ FRC reading cannot be over looked.

Small differences to FRC calculation have been shown to have a large impact on the error seen in modelled LCI [174]. Even large differences in FRC between two indirect N₂ washout devices have been described [165, 193, 219]. Poncin *et al.* 2016 completed a comparison of two N₂ MBW devices

and despite the washout gas being the same, they again found considerable differences between LCI and FRC results (Exhalyzer was 0.66L higher than the EasyOne Pro NDD device, $p < 0.0001$) [165].

They attributed the differences to gas delivery systems that changed breathing dynamics and measurement algorithms. Since the breathing dynamics in my study changed but only by a few breaths could mean that algorithm differences are the biggest explanation.

All of these studies suggest there may be an error in the Exhalyzer D system that overestimates FRC and in turn LCI above expected and traditional measures. It may be that the ExD/N₂ does not offer the most physiologically accurate results but as long as results are always compared to the same device, studies requiring an LCI output using an N₂ device should be able to use the Exhalyzer D. In fact a recent study of the drug Orkambi in children aged 6-11 showed LCI improved after 24 weeks of treatment using the Exhalyzer D [94].

In vitro validation studies have been performed using both Innocor and Exhalyzer D devices [119, 146] both showing promise as each device had repeatable and reproducible results when compared to a lung model. It seems that in vivo testing, particularly in CF where increased ventilation and longevity of the washout make the results more variable.

FRC measurement is crucially important, as it is the denominator in the calculation of LCI. If FRC is overestimated it will artefactually reduce LCI, potentially reducing its sensitivity to detect early lung disease. In section 4.4.1 I found that FRC from the ExD/N₂ was higher than the Inn/SF₆ but so was CEV. As LCI is a ratio of CEV and FRC it is likely that potential errors in FRC calculation are incorporated into the CEV measurement as well (particularly as part of the calculation of CEV and FRC come from the same washout end point). Therefore, derived LCI should be a true reflection of heterogeneity with an offset due to any potential errors.

Care should be taken with the interpretation of results, FRC results are not interchangeable and should be reported with the method it was derived from as is conventional for plethysmography or helium FRC. Due to the potential errors and differences between devices, only a long-term study using all of the devices would identify the most appropriate device for standardised use of FRC and in turn LCI.

4.5.3.2 Differences related to individuals' breathing patterns?

The main comparison of the modified Innocor device with the Exhalyzer D device highlights the discrepancies between the SF₆ and N₂ systems. One explanation for this result could be patient differences between tests. A simultaneous washout of both gases was attempted to eliminate these potential individual differences.

4.5.3.2.1 Simultaneous washout

Simultaneous washouts demonstrate that the difference between gases is not related simply to differences in the individual's performance when completing MBW's on different equipment, as breaths were the same for each gas and washout monitored at the same time. This study is important as it takes steps to unpick the causes of the washout technique differences and attempts to rule out one possible explanation for the differences seen. Breath by breath gas concentration taken from the Innocor device and Exhalyzer D confirmed that N₂ reached the 1/40th target concentration after SF₆ had. The lack of significance in CF ($p=0.06$) is likely due to small number and being underpowered.

As O₂ and CO₂ concentration are fairly constant with each inspiration and expiration during the washout, attempts were made to compare these results from each device. The breath-by-breath gas concentration for O₂ and CO₂ were not available since the devices do not compute it. Therefore the raw O₂ and CO₂ concentrations per millisecond were explored in a sample of patients (data not shown); the concentrations were similar in each device. Difficulties with the size of the raw data file and the ability to match the raw data from each device meant O₂ and CO₂ signals were not explored further. The O₂ signals were likely to be similar since it is the same oxigraph used to measure O₂ in each device. The infrared CO₂ technique in the ExD/N₂ and the photo acoustic detector in the Inn/SF₆ also seem to be comparable when the full raw data sample was plotted but as discussed this could not be compared further on a breath by breath level due to the size of the data file. With this in mind the difference could be the calculation of N₂ (please see section 4.5.3.5) as direct gas concentration detection of O₂ and CO₂ from both devices seems to be comparable and not likely to cause differences between the overall washout results. I consider that this experiment has ruled out one of the possible explanations for differences, an individual's breathing pattern. As the N₂ and SF₆ difference still persisted, further explanations were explored.

4.5.3.3 Tissue N₂

I next hypothesised that the difference could be caused by the contribution of tissue N₂ during the N₂ washout. To test this hypothesis, I extended the washout beyond its usual duration; this offered the opportunity to examine the decay curves of the gases. Figure 4.14 shows SF₆ decreasing down to zero whereas N₂ did not decrease much beyond the 2.5% target. Preliminary testing showed that increasing the extended washout period did not increase the likelihood of N₂ decreasing down to zero. When 100% O₂ is exposed to the gas sampling line, N₂ is calculated at zero so there is no lower limit of detection with the indirect measurement. This suggests that there genuinely is a N₂ signal present, most likely from the gas being generated by the host tissues, which prevents the signal from

reaching 0%. N₂ is not an inert tracer gas, there are stores and redox reactions where N₂ is produced within the body [172] and these may contribute to the washout signal (Figure 4.1). N₂ may contribute to the washout because 100% O₂ is used and this causes a concentration gradient from the lungs (abundant O₂) to the body (low O₂). Therefore N₂ travels from tissue and blood from a high concentration to a low concentration of N₂. It could also be that this lower plateau is the point where alveolar N₂ washout reaches a point where the flux from airway tissue is the same.

The MBW consensus statement and historical data suggest that in longer MBW tests i.e. more severe patients, FRC overestimation may be large as tissue N₂ may contribute to the signal [65, 101, 173]. Indeed using N₂ washout data, Nielsen *et al.* 2013, showed significant impact of increased N₂ concentration on the last part of the concentration curve; my observations that technique-related differences are more marked for LCI_{2.5} than LCI₅ are in support of this. They suggest at least 20% of the measured N₂ originated from the blood and tissues rather than the airways and the impact was variable particularly with increasing ventilation heterogeneity and dead space [174].

The results here suggest that N₂ release in the extended washout period is around 1.26%(0.09) in HC and 1.48%(0.12) in CF (the average minimal N₂ from the ExD recorded in the last two minutes after the 1/40th cut off). This is by no means a fool proof method as below 1/40th of the starting concentration the 'noise' and error in the system may increase the actual gas signal, this is particularly true as the N₂ concentration is calculated indirectly and this may exaggerate any errors. This method does not account for potential changes in N₂ release throughout the rest of the washout and reliance on the indirect calculation method. Only measuring N₂ directly would provide accurate information about N₂ release and it is likely to depend of duration of exposure to 100% O₂ breathing, body size and ratio of muscle to fat [220] as well as other variables such as gender, height, weight and cardiac output [174]. An abstract by Weiner *et al.* 2016 innovatively applied breath hold manoeuvres towards the end of MBW of SF₆ and N₂ with 100% O₂. They estimated the N₂ back diffusion by comparing changes in the washout trace of SF₆ and N₂. They found that N₂ concentration increased during a breath hold and attributed this to tissue N₂ and potential differences in cardiac output [221]. Discrepancies between exhaled breath concentrations in CF and HC participants have shown differences in known volatile organic compounds [222] and the nasal nitric oxide technique [223]. So, it may be that having CF and culturing different pathogens may increase the amount of additional N₂ contributing to the MBW signal. It could also be that the same 2 minute extended washout period here in CF is not long enough to clear N₂ to the same extent as seen in HC. These individual differences may all impact the rate of N₂ release.

One study offered a simple solution to adjusting the MBW for innate N₂ and harmonising N₂ results with SF₆, based on the minimal N₂ fraction achievable in healthy subjects [177]. They subtracted 1%

N₂ from each end tidal breath to adjust for the potential N₂ offset. They found that the adjusted N₂ washout achieved 1/40th (2.5%) 2 breaths sooner in one HC and 19 breaths sooner in one CF patient. The authors suggested that for SF₆ and N₂ washouts to be comparable, SF₆ washout should be extended to 1/66th (1.5%) compared to N₂ at 1/40th or N₂ washout should only be completed to 3.5%. As seen here, 1% N₂ may not be a large enough offset to account for N₂ contribution (average N₂ 1.26%(0.09) in HC and 1.48%(0.12) in CF seen here) however the authors highlight a plausible method to correct for N₂ should the known N₂ contribution be identified.

As direct measurement of tissue N₂ cannot be easily obtained through an inert gas washout and was not attempted in this thesis, a correction factor of the kind suggested by this paper has not been applied. It is noted however that a large difference in LCI values between gases could be due to final end point selection where the 1/40th cut off in N₂ undoubtedly captures N₂ washout but also N₂ tissue release. The study received some criticism however, since it assumed N₂ only had one time constant i.e. there was a constant release of N₂ as oppose to the known release at different rates and body organs and may be primarily mathematical rather than physiological [178, 224].

The accompanying commentary suggested that other discrepancies between N₂ and SF₆ should be explored more thoroughly, particularly that the Exhalyzer D system seems to be the common device where a large discrepancy occurs, so there may be an internal error, potentially relating to the indirect calculation of N₂ [178]. Indeed, the difference seen in my study is larger than reported previously and may even be larger than potential tissue N₂ contributions explain. The extended washout period however, does offer insight into innate contributions. The study by Yammine *et al.* 2015 does question whether use of an N₂ correction factor may now be required, particularly as N₂ washouts are being used more readily. It also asks whether the arbitrary 1/40th end point is still appropriate since it was only adopted because of the detection limits of historical devices. Sensors and devices are now much more accurate so it may be that for each test gas a specific end point could be defined. Caution should be utilised here as if MBW from N₂ does provide different information about ventilation heterogeneity to SF₆, then correcting the result to be in line with SF₆ would eliminate this information.

Clearly more investigation needs to take place with regard to the differing amounts of tissue N₂ that contribute to the washout signal. With increased use of MBW with N₂ as a tracer gas the understanding of tissue N₂ release is paramount for the validity of testing.

Tissue N₂ contribution is not easy to measure non-invasively and may well depend on cardiac output, blood volume and N₂ tissue stores (muscle and adipose) etc. [172] (Figure 4.1). Ultimately, care should be taken with the continuity of testing i.e. after a period of rest rather than exercise or not after a large meal as these may all affect the N₂ gas release; this will ensure realistic and repeatable

results. Some groups are working on a correction factor for N₂ but due to the complexity and unknown amount of the gas release it may take some time before this is put into practise. Olegard *et al.* 2005 suggest that a combination of washout and washin of N₂ could be used for more accurate estimations of FRC as this would eliminate the N₂ tissue factor [226]. The benefit may be that SF₆ and N₂ washouts could be compared making use of our historical data from SF₆ but this remains to be seen. Preliminary work on washin and washout MBW data was discarded due to the difficulties of the preferential filling and emptying during these phases.

Some would suggest that comparison of different methodologies is futile, particularly as lots of variables would need to be considered. It is also worth acknowledging that if results are different, what is the use in comparing them i.e. they will always be different. For now, knowledge of tissue N₂ contribution is essential but in general the amount is considered to be negligible and therefore no correction factor is recommended [101]. The current recommendation extends from historical data where N₂ release was measured in participants exposed to oxygen breathing [220]. The fact that MBW tests in children are only of short duration (2-3 minutes [141]), it is likely that the N₂ release is small. However as MBW testing sessions in adults and those with more severe disease can be in excess of 40 minutes [105], the N₂ release may be more culpable. Not only does this question the validity of MBW to measure ventilation heterogeneity (it may also be the degree of N₂ release being counted), it also questions the practicality and feasibility of MBW testing in these patients. There is certainly a contribution of tissue N₂ during a N₂ MBW, the amount is unknown but potential corrections should be considered, particularly as N₂ MBW may be utilised more readily.

4.5.3.3.1 Limitations with simultaneous washout testing

Raw gas concentrations

Using the raw breath-by-breath data may be a limitation of the study but these steps were taken to negate the difficulties with matching flow volume and signal synchronisation between devices.

Brunner *et al.* (1985) state that to get accurate readings during an N₂ washout, flow and gas concentration have to be synchronised using the delay time of the analyser [196]. Gas composition changes during a washout so the delay time also varies; Brunner *et al.* (1985) calculated volume and deadspace and found errors up to 70% if synchronisation was not performed. We ensured that individual devices were synchronised but matching two results from two separate devices was extremely challenging. By looking at raw data we have ensured we are not introducing more error with attempts to match data and that the discrepancies seen are real.

Preliminary attempts to calculate LCI using just one of the softwares available were abandoned as it was considered invalid due to different sample frequencies, recognition of the different gas sample

and differences in flow gas delay for each device. The software TestPoint (TestPoint, Capital Equipment Corp., Billerica, MA, USA), developed by Per Gustafsson and software using Matlab (R2006a; The Mathworks Inc., Natick, MA, USA), are able to calculate LCI from both N₂ and SF₆ but these were unavailable for use. Without calculating LCI results here it is not clear whether software algorithms add additional discrepancies between devices but the washout to 1/40th is slower in N₂ than SF₆ particularly in more severe disease.

Deadspace

The total deadspace seen with this set up is outlined in Figure 2.14 in the methodology section, the distance between devices and potential differences between the samples received at each device are also shown. The majority of studies looking at the effect of deadspace have been on infants and children [228, 229] as deadspace represents a larger proportion of their airways and considerations have to be made to ensure results are an accurate reflection of the airway [230]. Even in older children and adults, large equipment deadspace has been shown to increase LCI and FRC outcomes [174, 187, 189]. Measurement error with increased deadspace in calculated LCI was between 2-13% in CF children and the difference further exaggerated in more severe disease [174]. This may explain some of the differences seen between devices even though both follow published guidelines. The current recommendation from the MBW consensus statement is that deadspace should be <2ml.kg⁻¹ bodyweight in children, <1ml.kg⁻¹ in infants and an upper limit of 70ml in adults [101]. Deadspace considerations should also be incorporated into analysis algorithms to account for the gas sample that has not participated in gas exchange. One group has even suggested incorporating a deadspace correction factor [187]; whilst that is not appropriate here, it does highlight the importance of knowing the source of error and distinguishing deadspace differences from actual V_I changes in LCI. The deadspace here is only just above the recommended 70 ml in adults. The methodology here used raw values so deadspace was not accounted for, but the impact should be similar for both devices. A large deadspace above 70ml showed significant differences between LCI [187], but more investigation would be required to see if the deadspace volume utilised here would.

Deadspace and differences between sample collection points may clearly influence the gas signals and washout results. Whilst this set up is not recommended for clinical use, even with its limitations it offers a practical methodology to examine SF₆ and N₂ washout at the same time. As all participants were adults with tidal volumes above 450mls, the impact of a large deadspace is not as big as if the participants were children. Therefore this study offered proof that whilst individual differences may play a role in the discrepancy between SF₆ and N₂ washout when completed independently, removing them using a simultaneous washout does not eradicate differences seen.

N₂ and SF₆ washouts are inherently different from each other, higher for N₂. Tissue nitrogen is a key explanation for the difference between gases shown with extended washouts. Gas concentration, breath number and time to washout were used in an attempt to eradicate equipment differences when calculating LCI but the recorded results (gas concentration) were still collected from the separate devices.

The next step was to determine if 100% O₂ impacted the SF₆ washout and therefore influence the difference between N₂ and SF₆ MBW.

4.5.3.4 Respiratory Mass Spectrometry – Differences in gases

One question arising from this simultaneous method is whether the washout of SF₆ is different when completed using room air vs. 100% O₂. If the washout of SF₆ is affected by 100% O₂ then it may indicate that N₂ washout should be longer than SF₆ since it can't be washed out with room air and therefore the washout gas impacts the difference between MBW on SF₆ or N₂.

Using 100% O₂ as a washout gas does not impact the decay of 4% SF₆ compared to washout using room air (Figure 4.15). The addition of SF₆ to a standard N₂ washout also does not change washout breath number. This shows the comparisons and conclusions from the simultaneous washout section using raw data are valid. Clearly this study has a small sample size and using breath and time data alone is not ideal, but proof of principle is offered to assure findings from the simultaneous washout of SF₆ and N₂ at the same time. Only head to head data during the same washin and washout is likely to definitively assess differences in MBW [231] so this small study does take steps to combat this.

There were differences between the RMS and the other devices when completing a 1% SF₆ washout however (Figure 4.16). This included large variability of gas concentration detection and large signal to noise ratio towards the end of the washout (Figure 2.11 in the methodology section) suggesting that washout with 100% O₂ requires particular care, especially with low SF₆ gas concentrations. This is particularly important since in section 4.4.1 the largest differences between gases were seen in the tail of the washout. As CF participants were not recruited to this section because of the long washout times then the larger discrepancies in the washout tail may not even be noticeable here. Differences between gas properties will now be explored further.

4.5.3.4.1 Molecular properties of gases

The discrepancy between the two methods could be the physiological properties of the gases [232]. SF₆ is a dense gas when compared to N₂ (see section 4.1 and methods chapter) and it behaves differently within the lung compared to N₂. Differences in phase III slopes using different inert gases

in single breath washouts have been shown [197, 215]. The authors attribute the differences to different acinar diffusion points between tracer gases. Since we used multiple breath washout, it is likely that there is less impact of diffusion dependent inhomogeneity (which affect the phase III slope) as convection dependent inhomogeneity is dominant with consecutive breaths [47]. Therefore it could be that the impact of different tracer gases is smaller with MBW, although clearly it plays a roll, particularly towards the end of the washout where the poorly ventilated and or more peripheral airways are contributing.

Jensen *et al.* 2013 suggest that N₂ washout may be able to more accurately reflect the gas trapping and ventilation heterogeneity in the small peripheral lung units, in turn implying that SF₆ is unable to reach these areas (a partial washin, Figure 4.2). For example SF₆ may fail to fully washin to distal, obstructed lung units whereas N₂ is resident and is the main gas component of air and therefore a factor of normal breathing mechanics over a lifetime vs. just the washout period may mean that N₂ can more fully describe VI. Two older studies identified that density of gases within the lung influence the convection of oxygen and subsequent oxygen diffusion into the body [233, 234]. Less dense gas mixes increase the dispersion of oxygen to the periphery as there is more laminar flow compared to higher density gases that result in more turbulent flow [234]. However as the SF₆ diffusion convection front lays more peripherally than smaller gases [49] it seems unlikely that it is unable to disperse as far, unless there is obstruction in the more peripheral airways. Studies have alluded to different deposition of inhaled particles in the lung either because of particle size or extent of peripheral disease [181], but clearly drug particles may be larger than SF₆. The extent to which a partial or inefficient washin is accurate is unknown but researchers have suggested it may in fact increase an LCI signal due to algorithms determining different washout targets [235]. Therefore SF₆ washout may be larger than may be evident here so this hypothesis does not support my findings. The washin accuracy here was not directly assessed but all testing followed strict guidelines to ensure reliable tests. Only direct measurement of deposition would show the extent of washin.

To explore the differences in gas properties further, MBW derived from helium can be compared to SF₆. Helium washout was often completed simultaneously with the 4% SF₆ mass spectrometer data as it was part of the gas mix in this set up so many groups should have the data available. By looking at two gases with different diffusivities simultaneously, the theory was that it would reveal different causes of inhomogeneity, either obstruction (both gases) and or incomplete diffusion (SF₆) [236]. Despite having even larger molecular differences, one study did not show the same discrepancies seen in SF₆ vs. N₂ MBW results [118]. Variability in helium and SF₆ increased in disease but there was no bias of one gas over another. Unfortunately the data for helium washout was not shown in this study and other comparison data of helium and SF₆ is lacking. This would suggest that the difference

between SF₆ and Helium MBW results is small due to the underreporting of negative results [237] and the large number of sites that would have access to this data. Although one abstract did show the mean LCI was higher in helium compared to SF₆ but it had a small sample size [238]. LCI results from MBW studies using helium are in line with the results from healthy control SF₆ data in this thesis [231, 239] but further direct comparison studies would be required to be conclusive. This suggests that physiological properties of the tracer gas are not the cause of the discrepancy between SF₆ and N₂.

4.5.3.4.2 Washout gas

The above suggests it is not the tracer gas (N₂ or SF₆) but the washout gas (100% O₂ or room air) used in testing that affects the result and little is known about the impact of different washout gases as most inert gases are washed out with room air. Indeed 100% O₂ has been shown to change breathing dynamics but this may only be in children and infants [213] or very advanced disease when respiratory drive from hypercarbia has been blunted [240]. Possible mechanisms of breathing 100% O₂ include changing pulmonary vasoconstriction [241], ventilation perfusion [242] and breathing mechanics [213]. Hopkins *et al.* 2017 had 8 subjects complete 9 argon MBW's washed out by room air but using different test gases containing either 21%, 12.5% or 90% O₂. They found no differences in V_T, LCI, or time between different oxygen conditions but FRC was higher with lower O₂ (no differences when breathing 90% O₂) and deadspace was higher with higher O₂. The group considered the short duration of the MBW to be the main reason for a lack of difference to breath volumes and overall ventilation heterogeneity but implied that the results could not be extrapolated to those with lung disease since duration of washout is longer and O₂ breathing could have more of an impact (this is shown in section 4.4.1 & 4.4.2 where bigger differences exist in more severe disease). These MBW results were of a similar duration in HC so the lack of impact of increased O₂ may be similar. The mechanism behind discrepancies in FRC was unclear, particularly as washouts were of a short duration and breathing mechanics did not change. The authors touched on the small differences seen in physiological deadspace that can be associated with changes in cardiac output but the study had a small number of participants so further investigation on the influence to FRC should be explored.

Whilst this study looks at the impact of 100% O₂, it substitutes N₂ for argon (an inert tracer gas) so any differences seen when using N₂, which is not entirely inert, are negated. It also looks at the impact on O₂ on washin, rather than washout which are known to be different [231]. My study does offer insight into N₂ vs. SF₆ washout with 100% O₂, and also shows no quantifiable impact of breathing 100% O₂. However, more extensive research as shown by Hopkins *et al.* 2017 would offer a

more definitive answer to whether multiple breath washout is impacted by washout gas selection. The results from thesis suggest that breathing room air or 100% O₂ does not effect the washout of SF₆ (Figure 4.15) implying there is little impact to breathing dynamics. There is still a difference between N₂ and SF₆ washout however and this is difficult to extract since N₂ cannot be washed out with room air.

4.5.3.4.3 Viscosity

100% O₂ is more viscous than air and therefore could result in more turbulent flow within the lungs. Schilder *et al.* (1963) found that the effect of viscosity was greater than the effect of density at low lung volumes [232] therefore the washout gas may have more of an effect than the tracer. So, similarly to limitations to flow with SF₆ washin, the N₂ washout may be slower and this could result in the large discrepancies between MBW devices. The viscosity of gas may also impact the device's ability to accurately measure a gas mix due to the change of flow with a higher viscosity. This is particularly true with small concentrations of tracer gas at the end of the washout vs. the high abundance of washout gas. Therefore many devices apply a correction factor (Innocor device). Any errors in the measurement of a gas will be exaggerated with indirect calculation of N₂ therefore this could explain the difference seen between SF₆ and N₂ washout in the ExD.

The RMS has the capacity to detect gases up to 100% [194] so reading 100% O₂ should be possible. The errors seen here may be because 100% O₂ is viscous and may interfere with the readout capacity of the device or that 1% SF₆ and below is under the detection limit for the RMS.

It is known that gas compositions change throughout the washout i.e. decrease of the tracer gas and more abundance of the washout gas, and this can increase error in washout results [196]. It is also known that capillary length and gas mixing down the capillary can vary the result as well as different response times when switching between gas concentrations [194]. Both of these studies use high concentrations of O₂ on a mass spectrometer and there is no mention of the readout capacity using O₂. But historical MBW testing has predominantly been completed using SF₆ rather than N₂ so there may be an underlying reason for this.

Viscosity of the gas may be more likely to impact flow and FRC calculations since the sample uptake by the devices to measure gas concentration is smaller than the gas measured for flow calculation. However, all the devices used here have some side stream analysis so viscosity may always have an effect on the gas delay to the measuring device. With the underreporting of negative results it is difficult to tell why N₂ MBW using 100% O₂ has not been completed on the RMS. It is likely that viscosity of O₂ does not impact ultrasonic technology for flow measurement and indirect calculation of N₂ (ExD system) but it does when measuring gases directly (Innocor and RMS). This is because the ultrasonic device measures flow across a small diameter which has been shown to reduce the effects

of viscosity [196], whereas flow through a mesh pneumotach is likely to be affected by large pressure changes with high viscosity. This means that the issues experienced in this part of the testing may not be apparent with commercial N₂ MBW set-ups. If full MBW results are viscosity corrected [196] then MBW using N₂ on the RMS should be possible. Though not ideal, I negated the impact of using a highly viscous gas by only using raw gas concentration values. If this testing set up were to be utilised again then further calibration and testing for the impact of different gases would be required as well as more participants including those with CF.

The next step was to use a method to determine SF₆ and N₂ washout disparities that fully eradicated equipment differences.

4.5.3.5 Respiratory Mass Spectrometry – Do equipment differences explain N₂ and SF₆ differences?

Equipment for each technique is not the same, these include hardware differences such as flow detection (ultrasonic on the Exhalyzer D vs. flow detection using a pneumotach on the modified Innocor technique) and differences in gas concentration detection (direct measurement using photo acoustic spectrometry and laser diode absorption spectrometry for the modified Innocor device or laser diode spectrometry and infrared technology for the ExD/N₂ and indirect calculation of N₂ on the ExD/N₂ system). It also includes internal software algorithms such as how to calculate the end tidal gas concentration for each breath [243], different analysis software to calculate LCI (Spiroware for ExD/N₂ and SW software for the Inn/SF₆) and different ways to calculate the FRC [195]. Even small software changes on the same device have been shown to have a large impact on results [189]. The differences between devices can be seen in Table 4.1. With all of these things in mind it is unsurprising that the results are different between devices, particularly as SF₆ and N₂ are different gases.

RMS testing, washing out both SF₆ and N₂ on the same device at the same time, shows that equipment is not the main source of discrepancy between SF₆ and N₂ washouts. Though testing methodology encountered some errors (see Appendix 2), the final results offer proof of principle that longer N₂ washout than SF₆ is an inherent property of the test gas. The delta between N₂ and SF₆ is similar in RMS to the delta from N₂ and SF₆ calculated on the Exhalyzer D and modified Innocor device. This could imply that the raw values (gas concentration and breath number) are similar when calculated on the devices but once software algorithms are incorporated for the calculation of final LCI and FRC results then differences increase.

4.5.3.5.1 RMS limitations

RMS testing attempted to eliminate equipment differences as a potential source of disparity between SF₆ and N₂. The results for this section were imperfect and pushed the devices to their limits, there may also have been additional sources of error by attaching 3 devices in series. This section will explain the potential sources of error and justifications for the final methodology.

Calibration

Despite the response times of both SF₆ and N₂ on the RMS and other devices being similar the variability of the starting concentration of the gases mean that results may not be reliable enough to be definitive. Daily calibration of gases allows the fine-tuning of the signal and to set a realistic reference to ensure valid results [101, 244]. The RMS was calibrated to the required standards using a 4% SF₆ gas concentration that spanned the 1% concentration. A specialist 1% calibration gas mix would have been more accurate but due to lack of time for delivery, the expense and lack of utility after this small study, specific 1% SF₆ gas calibration was not ordered. The manufacturer implied that 1% SF₆ washout should be possible if the RMS sensitivity to SF₆ was manually increased and SF₆ signal to noise ratio was small. The Inn/SF₆ does not require specific calibration to SF₆ in routine calibration testing but as mentioned it was returned to factory 1% settings. The Inn/SF₆ device has previously been shown to be in line with the RMS [105], the differences in gas concentrations at the start of the washout meant that SF₆ measurement was not optimal but the tracking of each gas on a normalised log scale was a pragmatic solution.

Signal to Noise ratio

As 1% SF₆ is a low gas concentration, the ability to detect this is limited by variation in the measurement. The signal to noise ratio will get worse towards the end of the washout, as the concentration of the tracer gas gets smaller and the 100% O₂ (washout gas) gets larger. If the signal to noise ratio is too large, then the confidence in the accuracy of the measurement is lower. Figures 2.11 & 2.12 in the methods section show the RMS SF₆ and N₂ gas concentration data on the simple washout software, the SF₆ gas concentration has a higher signal to noise ratio than N₂. Although the impact of this signal to noise ratio cannot be directly measured, it does suggest that SF₆ detection at this low concentration is not optimal.

Encouragingly, Figure 4.17 shows similar washout decays of both SF₆ and N₂ obtained in figure 4.14 suggesting that a) the RMS is able to detect the gas concentrations used here and b) a similar pattern between SF₆ and N₂ i.e. SF₆ decreasing linearly and N₂ taking longer to decay and reaching a plateau. The larger difference between the dashed SF₆ lines may represent the increased signal to

noise ratio in the RMS device. The ExD/N₂ dashed line being after the RMS N₂ line may represent the introduction of software algorithms and potential discrepancies but this is speculative as it only increased by one breath in this example.

Raw gas concentrations

As with simultaneous testing using just the Ino and ExD, a limitation of the current study may be that raw gas concentration or breath data was used to draw final results from the RMS sections. These steps were taken to negate the difficulties with matching flow volume, signal synchronisation and dynamic adjustment for delay time between devices as well as LCI calculation not being possible on the RMS. The final RMS method calculated end tidal gas concentration from an average of each washout breath, which is a previously utilised method [101, 195]. The fact that gas differences still persisted offers proof of principle that equipment, specifically gas concentration measurement is not the main cause of the difference between N₂ and SF₆.

As I could not measure flow with our RMS I cannot eliminate that discrepancies between equipment algorithms and therefore FRC calculation would not contribute to the magnitude of the difference seen between LCI results. The extent of the difference in washout breath number between SF₆ and N₂ calculated on the RMS vs. the Ino and ExD was not significant. This implies that flow calculation may be source of larger error than gas concentration, but a smaller difference also existed in healthy controls in previous sections so in disease this may not be the case.

This testing was pragmatic and suggested proof of principle that N₂ inherently washes out after SF₆, considerably more refinement would be required if exploring the differences between SF₆ and N₂ using the RMS were to be completed.

4.5.4 Chapter Conclusions

N₂ and SF₆ washouts are inherently different, higher for N₂. A fundamental, non-linear bias in N₂ (even seen in HC individuals) should be considered when interpreting results. The Exhalyzer D results correlate well with MBW results from the modified Innocor yet the readings are significantly higher, particularly in CF, with larger discrepancies in advanced disease.

The variation between devices is in some cases larger than reported clinically relevant changes so are not dismissible. As the difference between SF₆ and N₂ is increased in more severe patients this questions the validity of MBW testing in this cohort as little is known about the impact of using one result over another for clinically relevant information.

The Exhalyzer D device provides a robust signal of gas mixing inhomogeneity; it is able to

differentiate between health and disease and it correlates with conventional Spirometry. Despite differences between the devices the SF₆ modified Innocor is also capable of these characteristics. The optimal device for clinical trials is likely to depend on the practicality of using N₂ as a tracer gas and or the requirement for physiological accuracy. Recent publications suggest harmonisation of manufacturer algorithms and in vivo MBW validation studies are required [165, 193]. This will help the standardisation of MBW and LCI as an outcome measure but results here show that equipment differences are not the only cause of discrepancies between N₂ and SF₆.

The difficulties with this RMS testing mean the results offer proof of principle rather than conclusive evidence that N₂ washes out after SF₆. This is also true because the sample size is small due to the complexities of testing. Multiple steps were performed to increase the accuracy of the result despite its limitations so confidence that research methodology was appropriate is high. Paired data analysis also increased the power of the result since SF₆ results were always lower than N₂ in each participant.

This thesis chapter attempted to quantify the four main explanations for differences between N₂ and SF₆ washouts and show why they are not comparable (Figure 4.18). The key explanation for the difference between gases is likely to be tissue N₂ contribution shown with extended washouts. Equipment difference cannot be eliminated but these results indicate that gas concentration measurement (O₂ and CO₂) is comparable between new commercially available devices. This questions the validity of flow, FRC calculation (higher FRC results to traditional body plethysmography), signal synchronization and how indirect calculation of gas concentration could explain the difference. In vitro validation studies have been performed in both Innocor and Exhalyzer D devices [119, 146] and although both SF₆ and N₂ devices follow published guidelines (MBW consensus statement) there are clearly differences in how results are derived. Alternatively, SF₆ could theoretically fail to reach poorly ventilated lung units or not wash in fully, and therefore fail to contribute to the signal.

The utility of the Exhalyzer D is supported by its use in a recent multi-center interventional study where LCI improved after 24 weeks of treatment [94]. But global acceptability of the device is not yet achieved since unexplained differences between devices, potentially due to inaccuracies in the calculation of FRC from the N₂ device are subject to scrutiny. These results indicate that further clinical studies are required to show whether there are any advantages of one device over another. Researchers may need to be pragmatic about requiring devices to be perfect before they are utilised and it is unlikely that only one methodology will be universally used. Software updates and learning over time will always occur (consider the new GLI reference equations for spirometry), there will may be an evolving competitive market of manufacturers and sometimes results, despite measuring

the same outcome, will be different (as shown here FRC from plethysmography or FRC from helium would always be reported as such).

Concerns for SF₆ usage are minimal when utilising the small gas concentrations applied here (modified Innocor technique). Even less SF₆ is being used in a new re-breathing system from Innovision (<http://www.innovision.dk/Products/Innocor.aspx>) that has shown promising results [111]. Therefore both N₂ and SF₆ washouts may well be utilised in parallel for MBW results in the future.

Different techniques may be optimal for different indications. Investigators undertaking basic physiology research may value signal purity so highly that any influences such as those described for N₂ are unacceptable. For others, there may be considerations, which balance this issue. As part of the European CF Society Clinical Trials Network's standardisation processes, sites' experience with N₂ washout and ease of access to 100% O₂ in all health care facilities influenced the decision to focus on the ExD/N₂ technique. In the clinical trial context, an outcome measure which is precise and provides a good signal to noise ratio is key; the use of ExD/N₂ to this end is supported by the data presented here.

4.5.5 Future Studies for the understanding of N₂ and SF₆ differences

Individual differences could be further explored using a lung model i.e. matching the flow and gas signals together with changes to breathing pattern, physiological deadspace and or viscosity etc. as this would offer simulation of these changing effects. However as shown, in vivo testing may be different to in vitro validation studies.

The Royal Brompton Hospital performs LCI measurements on the NDD EasyOne Pro device (N₂ washout) as part of the paediatric annual assessment. With the data here it would be interesting to perform a comparison to see if the EasyOne Pro is more in line with SF₆ washout and or if there are similar discrepancies to those seen by Poncin *et al.* 2016 [165].

With the raw data available from simultaneous washout of SF₆ and N₂ in this thesis, it would be valuable to calculate LCI and FRC from both N₂ and SF₆ on the analysis software Testpoint and or Matlab. This could definitively eliminate software algorithms as a source of error between the washout gases as the same analysis software applying the same algorithms could calculate the MBW results for both gases.

The RMS section was imperfect; whilst it is unlikely that centres will utilise the RMS going forward, the testing could be refined for tuning to the lower SF₆ concentrations or using a 4% washout gas. The incorporation of flow on the RMS would also negate having to use raw breath and gas

concentrations and start to eliminate comparison differences between equipment and analysis software.

Ultimately, SF₆ and N₂ results are different from each other so the main goal from this thesis should be to deliver that message and ensure appropriate safeguards are in place so that MBW results derived from different sources are not used interchangeably and clinical misinterpretation of results does not occur. Indices derived from MBW should be reported with the technique and test gas used as is conventional for FRC measurements; this will combat misinterpretation between LCI results.

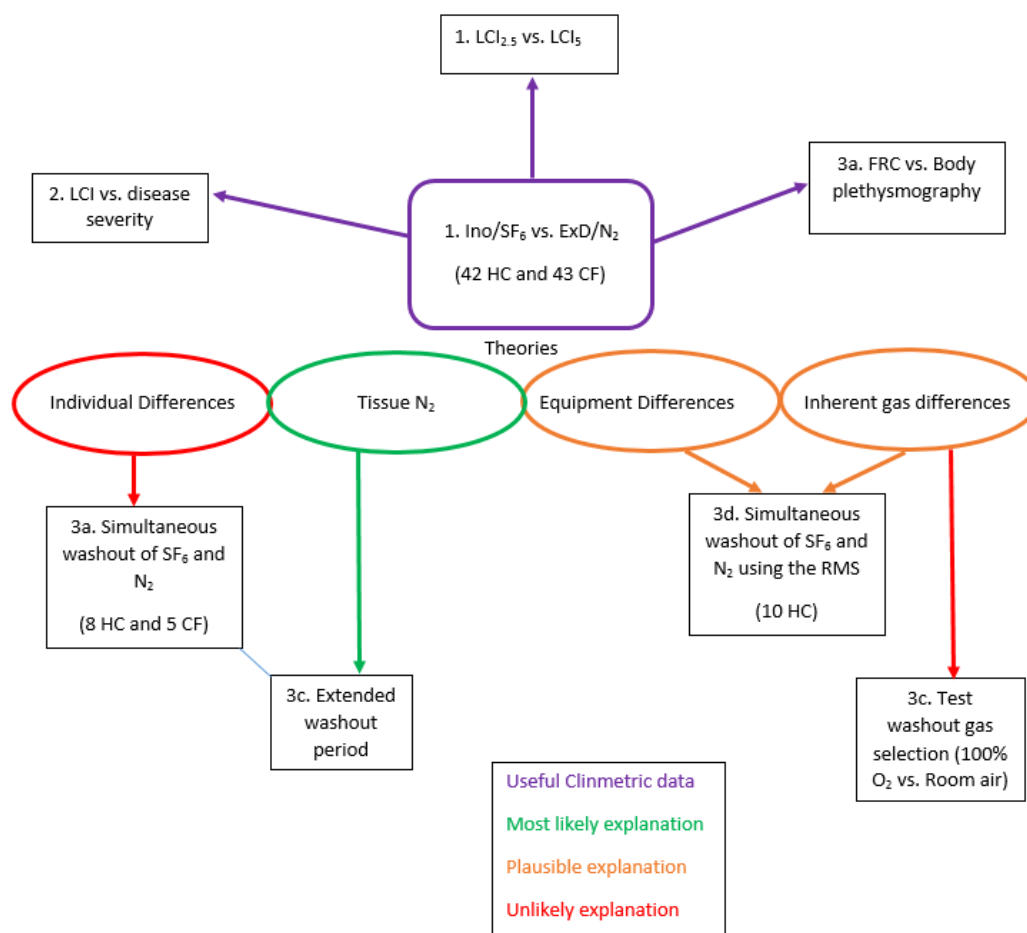


Figure 4.18 Summary of the different sections in SF₆ and N₂ chapter. Useful clinimetric data is highlighted in purple. The most likely explanation for the differences between SF₆ and N₂ is in green, plausible explanations in orange and unlikely explanations in red.

5 Chapter 5 - Exploring Lung Clearance Index as an outcome measure in the Cystic Fibrosis Gene Therapy Multi-dose Trial (MDT)

5.1 Introduction

This section describes the clinical trial 'A randomised, double-blind, placebo controlled Phase 2B clinical trial of repeated application of gene therapy in patients with cystic fibrosis' (MDT). The first section of this chapter will report the MDT study design and published results, the second will describe a further in-depth analysis of the MDT results from a physiological perspective, specifically LCI.

5.1.1 Gene Therapy and the Cystic Fibrosis Gene Therapy Consortium

Since the discovery of the CF disease causing mutation in 1989 [6], a plethora of studies have sought to provide gene therapy treatment to CF patients. CF is an attractive target for gene therapy as it results from a single gene defect. Essentially with the use of a vector (viral or non viral), a normal copy of DNA encoding the *CFTR* gene is transported into the host with the aim of producing functional protein [7]. Gene therapy has potentially very useful features over other experimental drugs and treatments targeting downstream manifestations of the disease: it is mutation class independent and therefore likely to be suitable for all patients; it targets the disease at the origin of the molecular defect [39]; and if resulting expression was of sufficiently long duration, treatment might only be required infrequently.

To measure success of gene therapy evidence of DNA delivery, successful transfection and presence and function of CFTR protein should ideally be found. CFTR DNA and mRNA are detected from sampling epithelial cells and measuring the presence of DNA/mRNA. Functional protein can be detected by nasal or bronchial epithelial potential difference measures, where correction of the chloride and sodium channels on the epithelial cell can be detected through exposure to different salt solutions and pharmacological agents [245] (<https://youtu.be/7JPuVbaLhTA>).

The first studies of CF gene therapy focussed on gene transfer efficacy in the nose and lung [246, 247] with viral vectors. The studies were single doses of gene therapy in small groups of CF patients yet there was evidence of return towards normal electrophysiology as measured by nasal potential difference and expression of normal CFTR. Repeated doses of the CFTR gene using a viral vector were then explored but there was variability amongst subjects and less correction with subsequent doses, perhaps due to increased immune responses [248]. A study using an adeno-associated viral vector to assess clinical benefit tested repeated application in 100 subjects [249]; whilst the delivery

was considered safe and well tolerated it did not achieve its primary endpoint of significant improvement in lung function over time.

This led researchers to explore the use of non viral vectors; in a double blind, placebo controlled trial, 12 patients received three doses of DNA and liposome or placebo to the nasal respiratory epithelial. 60% of the treated subjects tested positive for CFTR gene transfer as well as showing no apparent loss in efficacy of the dose on subsequent administration [250]. This highlights that unlike viral vectors; lipids do not provoke a strong immune response so again it may be considered a suitable vector.

After extensive preclinical optimisation of the lipid to improve gene transfer potency, safety and stability during aerosolisation, delivery of CFTR (plasmid pCF1-CFTR) using the lipid was completed in 1999 [251]. Delivery to the nose and lungs of CF subjects led to an overall ~25% correction of the CFTR chloride channel defect. However, the CFTR expression diminished within 1 week of administration. Furthermore, pCF1-CFTR/GL67A lung administration was associated with mild flu-like symptoms but they did resolve within 36 hours [251]. The studies above focussed on molecular rather than clinical outcomes so for gene therapy to be offered as a treatment, studies designed to detect clinical benefits would be required.

The UK Cystic Fibrosis Gene Therapy Consortium (GTC) was formed in 2001

(<http://www.cfgenetherapy.org.uk/index.html>); it pooled together scientists and clinicians in centres in London, Oxford and Edinburgh already completing gene therapy research. The GTC focused on providing proof of principle of gene therapy delivery and uptake but also on finding suitable assays for detection of *clinically relevant signals* within the CF population.

5.1.1.1 Gene delivery agents

Viral vectors, though very efficient at getting into the host cell, largely induce a host immune response which limits expression after repeat application [252]. Therefore non viral vectors, which cross the cell membrane by endocytosis [7], were utilised within the MDT.

5.1.1.2 GL67A

The liposome GL67A was chosen after *in vitro* and preclinical work confirming the vector was well tolerated with low toxicology in sheep studies [253]. GL67A is a cationic lipid combination (Genzyme Corp, Framingham, MA, USA), made from the following lipids (Avanti Polar Lipids (Alabama, USA) at molar ratios of 1:2:0.05:

- **GL67:** Cholest-5-en-3-ol (3 β)-, 3-[(3-aminopropyl)[4-[(3-aminopropyl) amino] butyl] carbamate]

- **DOPE:** 1,2-Dioleoyl-sn-Glycero-3-Phosphoethanolamine
- **DMPE-PEG5000:** 1,2-Dimyristoyl-sn-Glycero-3-Phosphoethanolamine-N-(¹-(3,5-bis(methoxy)phenyl)ethyl) (Polyethylene Glycol 5000)] (Ammonium Salt)

The lipid vector was considered the most optimal non-viral vector and was taken forward into further trials.

5.1.1.3 pGM169

pGM169 is a covalently closed circular double-stranded plasmid DNA molecule of 6549 base pairs purified from bacteria. Unlike previous CFTR plasmids, it was designed to have a long duration of expression with limited pro-inflammatory potential (VGXi, Houston, TX, USA). This was ensured by using a human elongation factor 1 alpha promoter, to “switch on” the CFTR gene, as well as the removal of unmethylated CpG dinucleotides, that are bacterially derived components on the surface of plasmid DNA likely to be responsible for flu-like symptoms upon dosing. By adapting the plasmid DNA in this way it was hoped that repeated administration would be achievable with minimal side effects.

5.1.1.4 Single Dose trial

A single dose safety study followed [254] (Sponsor’s Protocol Number: cro851; EudraCT reference: 2007-004050-85). The GL67A and pGM169 was shown to be safe when a single dose was administered to CF patients. Patients presented with mild influenza-like reactions after a single nebulised dose was given, but these reactions resolved within 30 hours. The reaction to the dose was in parallel to dose escalation and also further drops in lung function with larger doses. This led the 5ml dose to be taken forward into the multi-dose trial, as it was safe and the lowest dose in case of additive safety effects when the dose was to be given repeatedly. A single dose of the final active product (pGM169/GL67A) provided evidence of safety and efficacy of the formulation and enabled the selection of the appropriate dose with paracetamol being used as an adjuvant to minimize the risk of unblinding [254].

Although the single dose study was a safety trial, the group found vector specific DNA from lung brushings and biopsies at day 2 post nebulisation as well as some correction of the basic electrophysiological defect, via nasal and lower airway potential difference measurement. This highlights that not only can the gene be transferred across the lungs effectively but it can also influence the basic defect of CF.

5.1.1.5 Physiological Outcome Measures

A suitable outcome measure should be able to distinguish health from disease as well as detect changes after clinical interventions. An initial trial within the consortium measured changes in physiological, functional and structural biomarkers of CF after treatment for a pulmonary exacerbation [151, 255]. The group found that conventional lung function measurements showed significant improvements after intravenous antibiotics (IV's). They also explored novel outcome measures for identification of potential new biomarkers of treatment change and found decreased ventilation heterogeneity with LCI, decreased symptom scores and decreased systemic inflammation. The outcome measures selected here were able to show a change after treatment, therefore they were considered to be useful going forward into the monitoring of changes to treatment after gene therapy. This study was completed alongside a longitudinal Run-in study identifying long-term disease pattern in the CF population.

Starting in 2008, the Run-in study was longitudinal observational study of CF patients, assessing different outcome measures for the multi-dose trial (MDT). Patients in the run in study completed at least 4 visits a minimum of 3 months apart (visit 1-3 mainly in 2008, visit 4 in 2009, visit 5 in 2010 and visit 6 in 2011) at periods of clinical stability. The monitoring of patients over this period highlighted safety and efficacy assays to take forward into pilot studies of non-viral gene transfer agents, but also showed longitudinal changes of physiological outcome measures overtime. The candidate worked on these studies as a junior physiologist. Ideally these patients would have transitioned onto the MDT leading to a large database of baseline characteristics and then responses to gene therapy treatment, but funding issues introduced unanticipated complications.

The GTC then went on to complete the first phase IIb clinical trial to assess clinical benefit of repeated nebulised doses of non-viral CFTR gene therapy [181]. All of these steps are comprehensively covered by the group elsewhere [39, 256].

5.1.2 The role of the author on the MDT

For the first year of my PhD I was the main physiologist at the Royal Brompton Hospital and Imperial site with other members of the multi-disciplinary clinical team including a study coordinator, research nurses, research fellows and senior consultants. My roles were to perform clinical physiological testing (spirometry, LCI, gas transfer, cardio-pulmonary exercise testing and activity monitoring), writing standard operating procedures, designing and adapting the electronic data capture system, data handling, designing blinding procedures, recruitment, trial coordinator roles, multi centre communication and data analysis.

5.2 Hypotheses, Aims and Objectives

Hypothesis 1 – The MDT will provide the first clinical efficacy of repeated CFTR gene therapy

Aim 1 - Assess the clinical efficacy of the non-viral *CFTR* gene–liposome complex pGM169/ GL67A after repeated delivery to the airways

5.3 Methods

The methods here are reproduced with permission from the GTC.

5.3.1 Study Design and Study subjects

The MDT started in June 2012. It was a double blind placebo controlled trial where CF patients were randomly allocated on a 1:1 basis to either the active gene therapy product (5ml of cationic lipid complexed plasma DNA - pGM169/GL67A) or placebo (5ml of 0.9% saline) via nebuliser (Trundell Aeroclipse II device (London, ON, Canada).

136 patients were recruited to the trial, 116 of whom completed per protocol (receiving at least 9 doses). The dosing schedule was every 28 +/-5 days with an initial screening visit and two follow up visits (day 14 (FU1) and day 28 (FU2)) after the last dose. The MDT was the first large-scale trial designed to detect clinical benefit of non-viral gene transfer using relative changes of FEV₁% as the primary outcome.

The National Institute of Health Research (NIHR) Efficacy and Mechanism Evaluation (EME) programme funded the trial. The Gene Therapy Advisory Committee (GTAC -GTAC184) and Medicines and Healthcare products Regulatory Agency (MHRA - Clinical trials database EudraCT reference 2011-004761-33) approved the trial.

CF patients with an FEV₁ between 50-90% predicted at screening were initially recruited from the Royal Brompton Hospital in London and the Western General Hospital and Royal Hospital for Sick Children in Edinburgh. This range of patients were decided upon in that they were not too severe and obstructed that the gene could be delivered and that patients were not too mild so any potential gene delivery efficacy could be measured and detected. Recruitment involved screening hospital patient lists with knowledge of the full inclusion and exclusion criteria for the trial (Table 5.1). Once a list of eligible patients was identified, patients were contacted either by phone or in clinics to see if they would like to take part.

In January 2013 patient identification centres were opened in order to increase recruitment numbers, which involved contact with research coordinators at different hospitals across the country and these coordinators identifying suitable participants to come to either the London or Edinburgh sites to be screened. In June 2014 recruitment was closed with 136 patients enrolled.

Table 5.1 Inclusion and Exclusion criteria for the Multi dose Trial.

Inclusion Criteria	
1.	Cystic fibrosis confirmed by sweat testing or genetic analysis
2.	Males and Females aged 12 years and above
3.	Forced expiratory volume in the 1 st second (FEV ₁) between 50 & 90% predicted inclusive (GLI reference equations).
4.	Clinical stability at screening defined by: <ul style="list-style-type: none"> a. Not on any additional antibiotics (excluding routine, long-term treatments) for the previous 2 weeks b. No increase in symptoms such as change in sputum production/colour, increased wheeze or breathlessness over the previous 2 weeks c. No change in regular respiratory treatments over the previous 2 weeks d. If any of these apply, entry into the study can be deferred
5.	Prepared to take effective contraceptive precautions for the duration of their participation in the study and for 3 months thereafter (as stated in GTAC guidelines)
6.	If taking regular DNase (pulmozyme) is willing, and considered able by independent medical carers, to withhold treatment for 24 hours before and 24 hours after the gene therapy dose (nebulised doses only)
7.	Written informed consent obtained
8.	Permission to inform GP of participation in study
Exclusion Criteria	
1.	Infection with <i>Burkholderia cepacia</i> complex organisms, MRSA or <i>M. abscessus</i>
2.	Significant nasal pathology including polyps, clinically-significant rhinosinusitis, or recurrent severe epistaxis (nose bleeds) (nasal cohort only)
3.	Chloride secretory response on nasal PD of > 5 mV (nasal cohort only; will only be known after first measurement)
4.	Acute upper respiratory tract infection within the last 2 weeks (entry can be deferred)
5.	Previous spontaneous pneumothorax without pleurodesis (bronchoscopy subgroup only)
6.	Recurrent severe haemoptysis (bronchoscopy subgroup only)
7.	Current smoker
8.	Significant comorbidity including: <ul style="list-style-type: none"> a. Moderate/severe CF liver disease (varices or significant, sustained elevation of transaminases: ALT/ AST>100 IU/l) b. Significant renal impairment (serum creatinine > 150 micro mol/l) c. Significant coagulopathy (bronchoscopy group only)
9.	Receiving 2 nd line immunosuppressant drugs such as methotrexate, cyclosporine, intravenous immunoglobulin preparations
10.	Pregnant or breastfeeding

5.3.2 Randomisation

The participants were randomised on a 1:1 basis using an electronic database capture system (InForm™, version 4.6, Integrated Trial Management, Phase Forward Incorporated). Randomisation was stratified by site, age and FEV₁%.

5.3.3 Trial Triage

The 136 patients progressed through the trial after a successful screening pass i.e. fulfilling the inclusion and exclusion criteria; these patients were identified as the intention to treat group (ITT). A patient's first dose was completed within 28 days of the screening visit and after completing various outcome measures (to be discussed below). A 'safe to proceed' document was completed to ensure that the patient was well enough to continue with the dosing schedule i.e. their spirometry had not deteriorated by more than 10% and they had no symptoms that rendered them clinically unfit to dose. Each individual was scheduled to receive 12 doses, one every 28 days $\pm$ 5; patients were allowed to miss up to 3 doses if they were unwell or in exceptional circumstances where they were unable to attend. Patients who had at least 9 doses were designated the 'per protocol' (PP) group. At each visit a combination of outcome measures were performed and participants at the London site had the option to take part in subgroups for bronchoscopy and or nasal potential difference. Monitoring of patients and any adverse events were recorded immediately post dosing and throughout the study with the use of diary card. Usual treatment regime, any symptoms and the time until a subject's next pulmonary exacerbation were also recorded throughout the trial.

5.3.4 Outcome Measures

5.3.4.1 Primary outcome – FEV₁

The primary outcome of the trial was the relative change in FEV₁% after 12 doses of gene therapy or placebo. 120 completing patients were required to detect a relative improvement in FEV₁% predicted of 6% with 90% power. FEV₁% was completed at each visit. Pre dosing results were collated from screening (SCR) and pre dose 1 (DO1) visit. 12 doses were then completed and post-dosing results collated from both follow up visits (follow up 1, 14 days $\pm$ 2 days after dose 12 and follow up 2, 28 days $\pm$ 2 days after dose 12).

Percentage predicted scores for FEV₁ were calculated from the Global Lung Initiative (GLI reference ranges) [57]. Z scores were not utilised during the trial for ease of interpretation and the power calculation being based on percentage predicted. Problems with the extraction of data from the database mean that Z scores were also not utilised for additional analysis as part of this thesis. From the spirometry manoeuvre MEF₂₅₋₇₅, which represents the mid expiratory flows and therefore the smaller airways during the test, were also explored for comparison with other secondary outcomes and responses to treatment.

5.3.4.2 Secondary outcomes

Other safety and efficacy indices were also measured to assess patient responses to treatment and ensure any adverse reactions and any long-term effects of the gene therapy were thoroughly identified (Table 5.2).

Table 5.2 Outcome measures in the multi-dose trial.

Clinical efficacy	Safety
Relative change in lung function	Clinical examination
Lung Clearance Index	Transcutaneous oxygen saturations
Chest CT scan	Serum inflammatory markers
Exercise capacity (bike test)	Renal and hepatic function
Frequency of respiratory exacerbations	Gas transfer
Activity monitor	Immune response markers
Serum calprotectin	Endobronchial histology (bronchoscopy subgroup only)
Quality of Life Questionnaire	
Change in sputum weight	
Sputum culture and inflammatory markers	

5.3.4.2.1 Lung Clearance Index

Pre dosing LCI was obtained at the screening (SCR) and dose one (DO1) visit, which were averaged. Introduction visit (INT) LCI's were excluded from analysis as this visit was intended for new patients to the hospital and or to familiarise the patient with the new assays and to check for eligibility for the trial. INT data was used to assess whether there is a learning effect of LCI (Method development section). LCI was also collected at dose 4 (DO4), dose 7 (DO7), dose 10 (DO10) with post dose LCI calculated as the mean of follow up one (FU1) and follow up 2 visits (FU2).

The modified Innocor technique (Inn/SF₆) was used to gather LCI data (see methods section). At each time point, the patient completed an MBW test in triplicate in order to generate a mean LCI value.

Inclusion criteria for an acceptable test is show in Table 5.3. The data from the washout was exported and analyzed on "Simple Washout", an analysis program developed on IGOR Pro software within the GTC by Dr. Nicholas Bell as part of his MD. Extensive MDT standard operating procedures were developed and stringently followed for the collection of this data, an updated version of this SOP can be found here: http://www.mbwtraining.com/UKCFGTC_MBW_SOPsept2010.pdf.

All of the MDT LCI data at the London site were analysed independently by the candidate. Cross site comparison and data checks were performed with Dr David Armstrong (clinical research fellow at the Edinburgh clinical site) and inter-site quality agreement was good; 854 LCIs were performed in the trial, 118 analysed by both assessors with a 96.6% agreement [211]. LCI was a secondary

outcome within the MDT and the published data only reported the basic results. Therefore more detail on LCI baseline characteristics, differences between groups and LCI subsets will be explored here.

Table 5.3 Acceptability criteria for LCI inclusion.

Acceptability criteria of LCI traces for the MDT
<p>Daily calibration of flow and gas within limits of agreement</p> <p>Wash in difference between inspiratory and expiratory tracer gas is <0.03%</p> <p>Clear disconnection quality</p> <p><u>Traces show:</u></p> <p>A degree of uniformity for tidal breathing</p> <p>Absence of leaks or irregular spikes</p> <p>Minimal difference between 3 tests for FRC and LCI is <10%</p> <p>At wash out, concentration of tracer gas is <0.02%</p> <p>At least 2 of 3 tests deemed as adequate</p>

5.3.4.2.2 Chest CT Scan

As described in the methods section a CT scan images the chest to identify specific areas where lung disease is present. A full chest CT scan was completed at SCR (pre dosing) then at FU2 (post dosing). The main parameters from the chest CT scan utilised here are large and small mucus plugs as well as extent of gas trapping. The two scans were independently scored and results compared pre and post trial.

5.3.4.2.3 Frequency of respiratory exacerbations

Chronic progression of CF lung disease is characterised by intermittent episodes of acute worsening of symptoms and after a pulmonary exacerbation lung function is not likely to return to baseline [257]. By monitoring the frequency of exacerbations over a long period of time treatment response can be detected since exacerbation number may fall therefore potentially reducing the requirement of clinical interventions. Respiratory exacerbations were counted and graded as adverse events during the trial and the number of oral or IV antibiotics according to the treatment of the exacerbation was recorded.

5.3.4.2.4 Quality of Life Questionnaire

The CFQR QoL is a validated questionnaire for CF patients [81]; it provides a measure of patient reported outcomes including responses to respiratory health status and symptoms over the past two

weeks. The respiratory domain has shown the most utility with regard to treatment interventions therefore this will be utilised here. The CFQR QoL was completed at SCR, throughout the study then at FU2 but the pre and post perception of respiratory status will be explored here.

5.3.4.2.5 Confirmation of gene transfer

Quantification and the presence of DNA and mRNA were assessed by analysis of nasal and bronchial brushings. Functional protein was assessed by electrophysiological testing of the nasal and bronchial mucosa via stimulation of the epithelial with salt solutions. Please see the supplementary index from the main MDT for more details [181].

5.3.5 Dose administration

Both study team and study subjects were blinded to the treatment arm for each participant. The unblinded hospital pharmacy staff at each clinical site prepared the active or placebo dose and decanted it into masked AeroEclipse II breath actuated nebuliser pots and opaque nasal spray bottles. Before dosing, each patient completed the outcome measures for the assigned visit and their safety to proceed was also assessed (Table 5.4).

Providing the subject fulfilled the entire “safe to proceed” criteria the dosing regimen began. To prevent bronchoconstriction related to the hypotonicity of the gene therapy product, each subject received 200-400mcg of inhaled salbutamol then started the 40 minutes of breath-actuated inhalation of the study dose. This was completed with cycles of 3 minutes breathing the product through the nebuliser and mouthpiece then 2 minutes rest (Figure 5.1). Immediately post dosing and at 6 hours post-dose, 1g of paracetamol was given to reduce inflammatory symptoms seen previously in the pilot study [254] and to maintain the blind.

Table 5.4 Summary of study assessments

Summary of Study Assessments																										
Week								0	4		8	12	16	20	24	28	32	36	40	44	46	48			49	
Visit name	E&C	Intro	Screen	nPD1 ^E	nPD2 ^F	nPD3 ^E	Bronch ^H	dose1	D2a	dose2	D2b	dose3	D2c	dose4	dose5	dose6	dose7	dose8	dose9	dose10	dose11	dose12	F/U1	F/U2	Bronch	
CONSENT/ HISTORY/ EXAMINATION																										
Informed Consent	X ^A	X ^A	X ^A				X ^I																			
Medical History	X	X	X				X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	
Physical Examination	X	X	X				X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	
Quality of Life Questionnaire (CFQ-UK)			X											X			X		X				X	X		
Vital signs ^I	X	X	X				X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	
Pulse oximetry	X	X	X				X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	
Prior medications	X ^A	X ^A	X ^A																							
Concomitant medications			X				X	X		X		X		X	X	X	X	X	X	X	X	X	X	X	X	
Check diary comments and reissue diary								X		X		X		X	X	X	X	X	X	X	X	X	X	X	X	
Adverse events							X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	
Download Piko-6 device			X ^K				X	X		X		X		X	X	X	X	X	X	X	X	X	X	X	X	
Sweat test ^R	X ^R	X ^R	X ^R											X	X	X	X	X	X	X	X	X	X	X	X	
SAMPLES ^J																										
Blood sampling	X	X	X					X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	
Urine sampling	X	X	X					X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	
Sputum ^N		X ^{L,P}	X ^{L,P}					X ^L	X	X ^N	X		X					X	X ^P	X ^P			X ^M	X ^M		
Sputum for NTM	X ^M	X ^M	X ^M					X ^M															X ^M	X ^M		
24 hour sputum weight		X	X					X	X	X	X	X	X	X				X					X	X		
INVESTIGATIONS ^J																										
Lung Clearance Index		X ^L	X ^L					X ^L						X					X				X	X		
Exercise Bike test			X ^M					X ^M						X ^M	X ^M				X					X		
Activity monitor			X													X							X			
CT scan			X											X ^F										X		
Gas transfer			X						X		X		X	X			X							X		
Bronchial blood flow measurement ^{LS}			X											X										X		
Spirometry	X	X	X				X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	
DOSING																										
Randomisation ^{LS}								X																		
Study dose administration								X		X		X		X	X	X	X	X	X	X	X	X				
NASAL/ BRONCH SUBGROUPS ONLY																										
Nasal potential difference	X ^C	X ^C	X ^C	X ^C	X ^C	X ^C						X ^B	X ^A		X ^B	X		X ^B	X ^A			X ^L	X ^L	X ^L		
Nasal brushing			X ^B			X ^B																		X ^M	X ^M	
Bronchoscopy							X																		X	
PRE-DISCHARGE ADMIN																										
Instructions re home spirometry (Piko-6)		X	X					X		X		X		X	X	X	X	X	X	X	X	X	X	X		
Instructions re use of symptom diary		X	X					X		X		X		X	X	X	X	X	X	X	X	X	X	X		
Schedule next study appointment		X	X					X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X		
Superscripts																										
A- first point of contact only																										
B- to be conducted within week prior to visit																										
C- 3 occasions only																										
D- one occasion only with final nPD																										
E- may occur before Screening																										
F- limited cut																										
G- if available; timing may vary																										
H- bronchoscopy-specific consent only																										
J- on dosing visits, must be pre-dose																										
K- Patients who have undergone Intro visit only																										
L- 2 time points only																										
M- one time point only																										
N- for intensively monitored cohort only																										
P- Can induce at these timepoints; required once pre and once post dosing																										
Q- if no sputum available, cough swab to be performed																										
R- one time point pre-dosing if required																										
S- Optional at day14 post dosing																										
Subgroups in grey areas:																										
Intensive cohort, new subjects,																										
nasal/ bronch cohorts																										

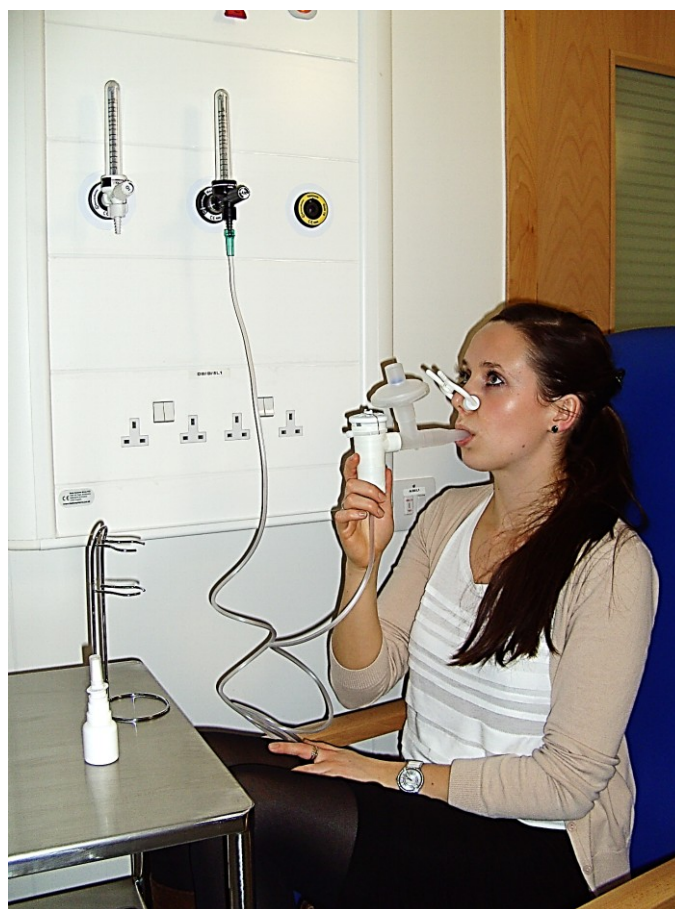


Figure 5.1 Demonstration of the lung dose using the blinded AeroEclipse II breath actuated nebuliser.

5.3.6 Data Analysis and Statistics

Data recorded at the visits was documented in individual patient case record files; it was then entered onto the electronic data capture system InForm with the Imperial Clinical Trials Unit (ICTU) verifying source data.

Once the data entry was complete the data was locked and downloaded for analysis by ICTU, a trial statistician and senior researchers as part of the gene therapy consortium. Statistics were performed as part of the statistical analysis plan approved by a trial steering committee.

An ANCOVA model was used adjusting for sex, study site, baseline FEV₁%, treatment allocation and subgroup, to compare between groups. IBM SPSS Statistics version 22.0 (IBM, Armonk, NY, USA) and GraphPad Prism version 5c and 6 (GraphPad software, La Jolla, CA, USA) were the statistical software packages utilised.

5.4 Results

Primary and secondary outcome measures relevant to this thesis were shown as reported in the MDT publication [181].

5.4.1 Patient groups and Demographics

A total of 191 CF patients were recruited to the study across both clinical sites. Figure 5.2 shows patient progression across the trial. After some screen failure and patient attrition due to exclusion criteria, clinical instability or patients withdrawing consent, 136 patients completed at least one dose of study drug (ITT group). The final data set (Per Protocol (PP) group) consisted of 116 CF patients all completing at least 9 doses. In this thesis, only the PP group will be referred to for analysis. After randomisation at the screening visit, 54 patients were allocated to the placebo arm and 62 to the treatment arm.

Baseline demographics for the PP group are described in Table 5.5.

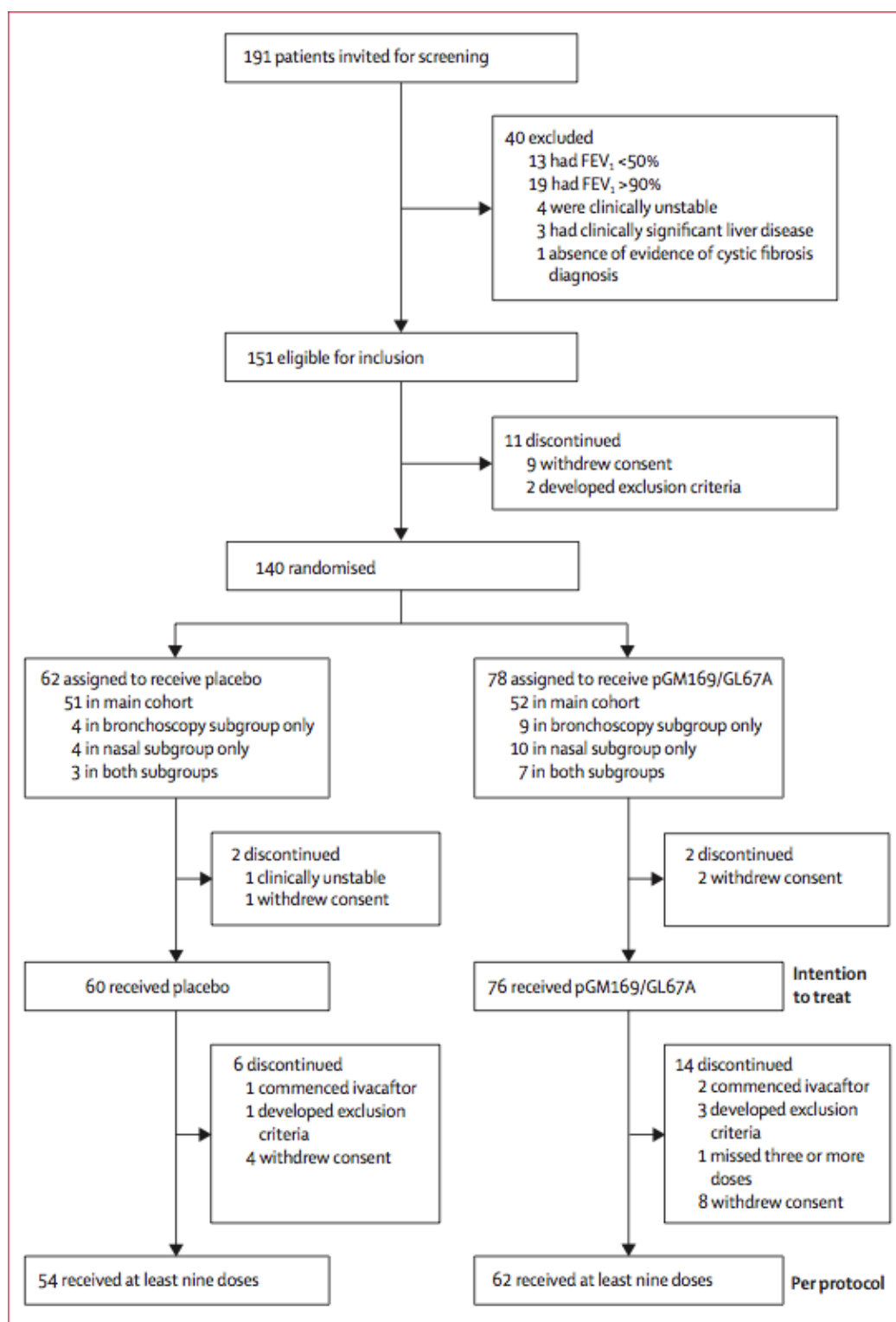


Figure 5.2 Consort diagram showing patient progression throughout the trial. 116 patients formed the “Per protocol” group receiving at least 9 doses throughout the trial.

Table 5.5 Baseline characteristics of the 116 participants in the MDT (as per published data by Alton *et al.* 2015).

	Placebo group (n=54)	pGM169/GL67A group (n=62)
Age (years)	26.0 (13.0)	23.6 (10.8)
<18 years old	17 (31%)	23 (37%)
≥18 years old	37 (69%)	39 (63%)
Sex		
Female	25 (46%)	31 (50%)
Male	29 (54%)	31 (50%)
Centre distribution number		
Edinburgh	24 (44%)	22 (35%)
London	30 (56%)	40 (65%)
Height (cm)	165.0 (10.6)	163.6 (10.9)
Weight (kg)	61.6 (15.6)	61.0 (15.7)
FEV ₁ (% predicted)	69.0 (9.9)	69.9 (11.1)
Body-mass index (kg/m ²)	22.4 (4.4)	22.4 (4.5)
Mutation class		
Phe508del/Phe508del	26 (48%)	31 (50%)
Phe508del/class 1-6	22 (41%)	23 (37%)
Not Phe508del/class 1	1 (2%)	3 (5%)
Heterozygous/homozygous class 3-6	2 (4%)	2 (3%)
Phe508del/unknown class	3 (6%)	3 (5%)

Data are mean (SD) or n (%), unless otherwise indicated.

5.4.2 Primary Outcome

Results for the primary outcome (relative % change in FEV₁% across the trial) were available from 98% of the trial patients; two patients were unable to provide a follow up FEV₁% result due to i) a pneumothorax related to an indwelling central line replacement and ii) withdrawal from the trial due to time commitments. The placebo patients showed a mean relative change of -4% (95% CI -6.6 to -1.4) in FEV₁% over the course of the trial whereas the active group changed by -0.4% (-2.8 to 2.1) (p=0.046) (Figure 5.3, graph A). Serial time point graph (Figure 5.3, graph) demonstrates that the Active treatment group appear to have a better preserved FEV₁ in the earlier stages of the trial (Dose 4). Figure 5.3, graphic B and C represent individual patient responses (relative percentage change in FEV₁% from pre to post) to treatment in both the active and placebo groups.

When splitting the groups by severity (<70% and >70% FEV₁ from baseline), though not significant there was a clear distinction in response to treatment in those with more severe disease than those with milder disease (Figure 5.4). In more severe patients the treatment effect was 6.4% (95% CI 0.8-12.1) and in milder patients the treatment effect was 0.2% (-4.6-4.9) p=0.065.

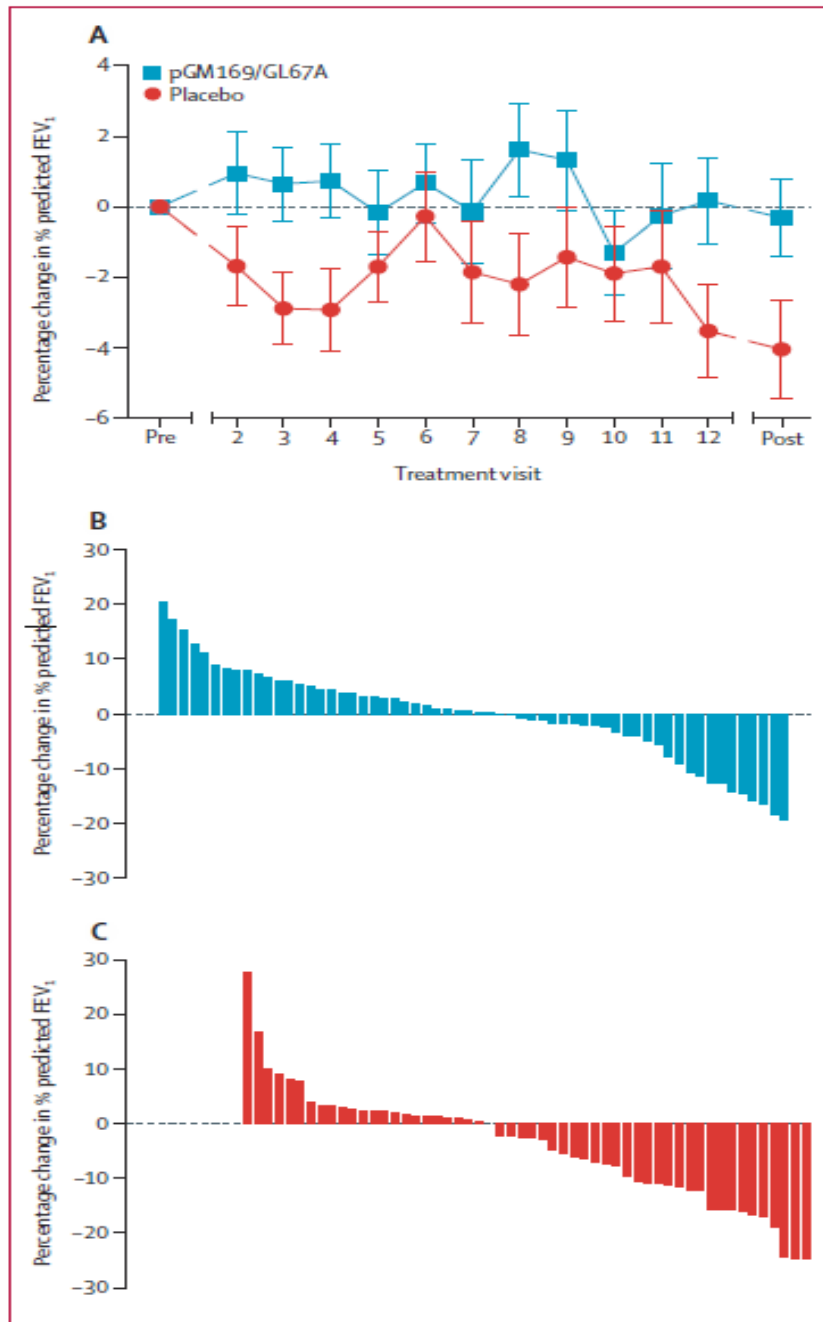


Figure 5.3 Primary outcome results. Graphic A shows the FEV₁% over the time course of the trial (Placebo red, Active blue). Graphic B and C show the individual patient responses pre and post trial (Placebo red, Active blue). Taken from the MDT published results [181].

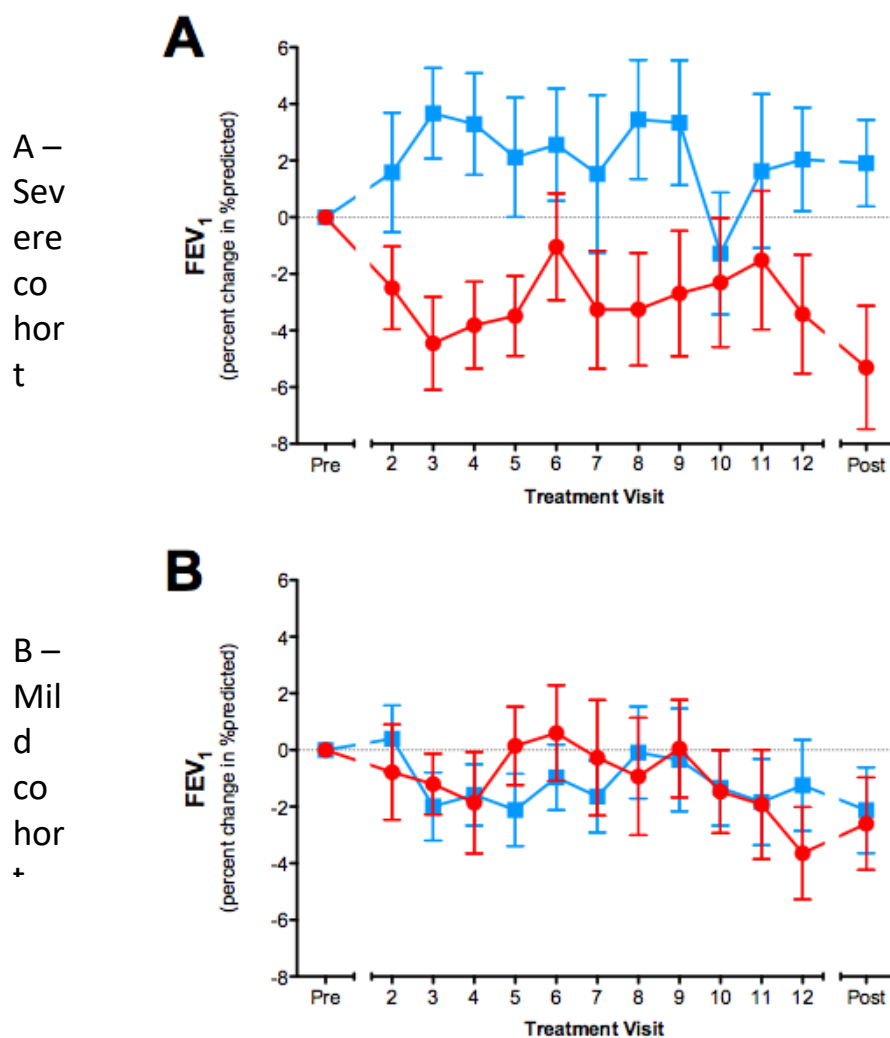


Figure 5.4 Time course of the primary outcome response stratified by baseline FEV₁% at trial entry. Red circles indicate the Placebo group and blue squares the Active group. Error bars indicate standard error of the mean. A. Baseline FEV₁% between 50-70% predicted. B. Baseline FEV₁% between 70-90% predicted. A. The treatment effect was 6.4% (95% CI 0.8-12.1), B. The treatment effect was 0.2% (-4.6-4.9). Interaction p=0.065, ns.

5.4.3 Major Secondary Outcomes

LCI data was available for 51/54 patients in placebo and 59/62 in active (data excluded at analysis or patient drop outs). Absolute delta change in LCI over the trial was not significantly different between groups (p=0.187) (Figure 5.5 and Table 5.6).

Over the course of the trial, both groups demonstrated a significant deterioration in LCI: Placebo pre 11.12 ± 2.47 , post 11.64 ± 2.11 ($p=0.007$). Active pre 10.37 ± 2.4 , post 10.77 ± 2.39 ($p=0.002$). The change from SCR to FU is within LCI variability so it could reflect regression to the mean. Serial time point graph (Figure 5.5) demonstrates that the Active treatment group appear to have a better preserved LCI in the earlier stages of the trial (Dose 4).

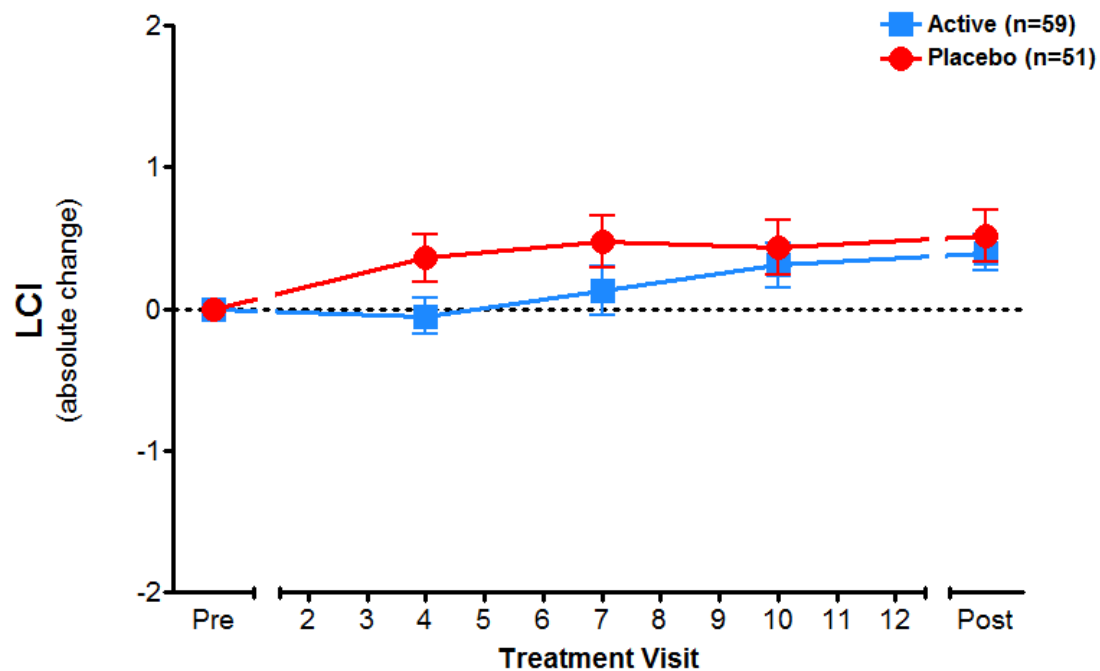


Figure 5.5 Time course of the response of LCI to either Placebo or Active. 'Pre' and 'Post' indicate the mean of two measurements at the respective time points. Error bars indicate standard error of the mean.

Table 5.6 shows treatment effects reported in the trial, for outcome measures referred to as part of this thesis. Data is reported as mean (95% confidence interval). To allow results from different endpoints to be plotted on a common scale, the estimated treatment effects were standardised and presented as multiples of the underlying SD (standardised treatment effect). Only FEV₁% and gas trapping reach statistical significance over the whole trial.

Table 5.6 Absolute and standardised treatment effects of the major primary and secondary outcomes in the trial (FEV₁%, LCI, MEF₂₅₋₇₅, CT parameters (Small mucus plugs, Large mucus plugs, gas trapping) and the respiratory domain from the CFQR QoL.

Outcome Measure	Absolute treatment effect (range)	Standardised treatment effect (range)	Adjusted p value
FEV₁%	3.66 (0.07 to 7.25)	0.39 (0.01 to 0.77)	0.046
LCI	-0.28 (-0.71 to 0.14)	0.26 (-0.13 to 0.66)	0.187
MEF₂₅₋₇₅	0.07 (-0.08 to 0.22)	0.18 (-0.21 to 0.56)	0.362
CT – Small mucus plugs	-0.07 (-0.15 to 0.00)	0.37 (-0.02 to 0.75)	0.061
CT – Large mucus plugs	-0.03 (-0.1 to 0.04)	0.18 (-0.21 to 0.56)	0.369
CT – Gas trapping	-3.49 (-6.96 to -0.03)	0.39 (0.00 to 0.77)	0.048
CFQR QoL – Respiratory domain	2.08 (-3.06 to 7.22)	0.16 (-0.23 to 0.54)	0.425

Vector specific DNA could be detected from bronchial brushings but subsequent mRNA was not quantifiable in either group, showing that DNA transcription was not recorded.

Corrected protein was measured by potential difference; nasal potential difference showed no difference pre and post trial however the bronchial PD was significantly different, the treatment group having PD responses in the negative direction that is considered the normal range [181]. These subgroups were not powered to detect treatment benefit.

In summary, the trial reached its primary end-point. Although not shown here, treatment effect analysis for all of the secondary outcomes favoured the active group (Forest Plots featured in the supplementary material of the main trial paper [181]) and there were trends for treatment changes when splitting the group by severity, more change in those with the more severe baseline (Figure 5.4). The secondary outcome measures shown here were not significantly different between groups, apart from reduced gas trapping on the CT scores post treatment (Table 5.6). LCI is said to be more sensitive than spirometry [74] yet there was no significant response to gene therapy treatment seen here, this will be explored further below.

5.5 Further MDT analysis undertaken in this thesis

The MDT will now be explored from an LCI perspective.

5.5.1 Research questions

Post hoc analysis of the MDT results found that the only variable associated with improvement in FEV₁% in the active group was baseline severity (Figure 5.4). Those with a lower baseline FEV₁% (50 to 70%) had a larger improvement compared to those with less severe baseline (70 to 90%). The explanation considered most likely by the GTC was that in this more severe group, a greater pGM169/GL67A dose was deposited in the larger airways due to obstruction in the peripheries and therefore widespread dispersion of the dose was not possible [181]. We know that FEV₁ mainly measures proximal large airways. If this theory is correct, more CFTR was deposited and expressed in these large airways (functional protein is present and therefore correct sodium and chloride transport takes place leading to airway clearance) and FEV₁ showed improvement, as the large airways were cleared of mucus.

LCI is a whole lung measure, but it is particularly sensitive to inhomogeneity in the small airways [74]. If the dose deposition theory is correct then although the larger airways may be clearer, the small distal airways will not have received a similar level of therapy due to obstruction and therefore this group may not demonstrate an LCI change; the converse may be the case in the milder group, if more distal- but possibly more dispersed- deposition may lead to an LCI signal even in the absence of one in FEV₁.

The literature tells us that LCI is a sensitive outcome measure [105], more sensitive than spirometry [74] and is responsive to treatment [38]. This was not the case during the MDT.

5.6 Hypotheses, Aims and Objectives

Hypothesis 1 - The absence of a treatment effect on LCI reflects either its limited utility in advanced disease or a lack of peripheral deposition in these patients.

Aim 1 – More fully understand the LCI signal from the MDT

Objective 1 - Assess baseline characteristics of physiological outcome measures to ensure LCI results for the treatment groups were in line with previously reported literature.

Objective 2 - Look at delta changes throughout the trial and the influence of disease severity on that change.

- a. Does a subject's baseline measurements relate to their delta change over trial?
- b. Do delta changes in LCI and FEV₁% compare?

Objective 3 - As patients were stable at the beginning of the trial and there seems to be an initial separation of FEV₁% and LCI at Dose 4, explore the possibility that LCI is able to read a signal in more stable patients where there is less noise from exacerbations etc. at this earlier time point.

Objective 4 - Compare LCI with FEV₁% and other outcome measures to assess responses to treatment.

a. Do other small airways outcome measures (MEF₂₅₋₇₅ and CT parameters) correlate better with FEV₁ or LCI?

Objective 5 - Do CFQR QoL results relate to FEV₁% and LCI results and add any information about an MCID?

Objective 6 - Review the subset of patients who have prior longitudinal FEV₁% and LCI data from the Run in study and assess if the trial changed the trajectory of the outcome measures.

5.7 Analysis methods for this thesis

There were no adaptations to the MDT methodology for use of the data for this thesis, however different analyses were completed to assess the LCI results from the trial more thoroughly.

Demographic data were explored for the whole CF group and between treatment groups to ensure that both Active and Placebo groups were evenly distributed and this could not be the explanation for any reported differences seen.

As initial trial results were manipulated and in some cases split into subgroups, the ANCOVA model was not deemed appropriate and therefore not adopted for trial results calculated for this thesis. Group differences were analysed using paired t-tests and ANOVA analysis.

Relationships between continuous and rank ordered variables were assessed using correlation coefficients and linear regression with r^2 and Chi squared used to assess patterns of change. Slopes were compared using the compare slopes function in GraphPad; using the null hypothesis that the slopes are all identical it tests if slopes were significantly different in a similar method to ANCOVA. All analysis was completed on the same software packages utilised in the trial. All of the outcome measures additionally explored in this thesis are physiological measures all shown to be associated with LCI [74, 75, 258]. These include: FEV₁%, LCI, MEF₂₅₋₇₅, CT markers (small mucus plugs, large mucus plugs, gas trapping) and CFQR QoL results. In the event of multiple comparisons I considered a p value of <0.01 to be significant.

5.8 Results

5.8.1 Baseline characteristics from an LCI perspective

Since LCI did not improve significantly with active treatment, I looked at baseline characteristics to establish if LCI correlated with other outcome measures as would be expected. The pre dosing results were taken from results collated from the SCR and DO1 visit i.e. a mean of data if collected at both. For this section, both active and placebo groups were merged to establish whether our PP group were characteristic of a CF population seen in other studies.

The mean $\pm$ SD pre dosing LCI was 10.78 ± 2.45 , and FRC was 2.20 ± 0.62 .

There were no differences when splitting the population by sex, clinic site, genotype classes or body mass index (underweight (<18.5 , $n=18$), normal ($18.5-24.9$, $n=69$), overweight ($25-29.9$, $n=19$) and obese (>30 , $n=6$)). LCI was lower in <18 years ($n=39$) compared to adults ($n=74$) (10.05 ± 2.1 vs. 11.17 ± 2.53 $p<0.02$).

The pre dosing LCI data was then compared between the FEV₁ severity groups assessed in the previous analysis (50-70%, mean (SD) $60.8\% \pm 5.4$, $n=58$) and 70-90%, $78.4\% \pm 5.6$, $n=55$). As expected, LCI was significantly higher in the more severe group (12.0 ± 2.3 vs. 9.5 ± 1.9 , $p<0.001$). There was a significant correlation between FEV₁% and LCI in the group as a whole $p<0.01$; this was also observed in the severe group $r=-0.4$, $p<0.01$ but not in the mild group $r=-0.2$, $p=ns$ (Figure 5.6) showing the added benefit of LCI in mild disease since LCI is more abnormal than FEV₁%.

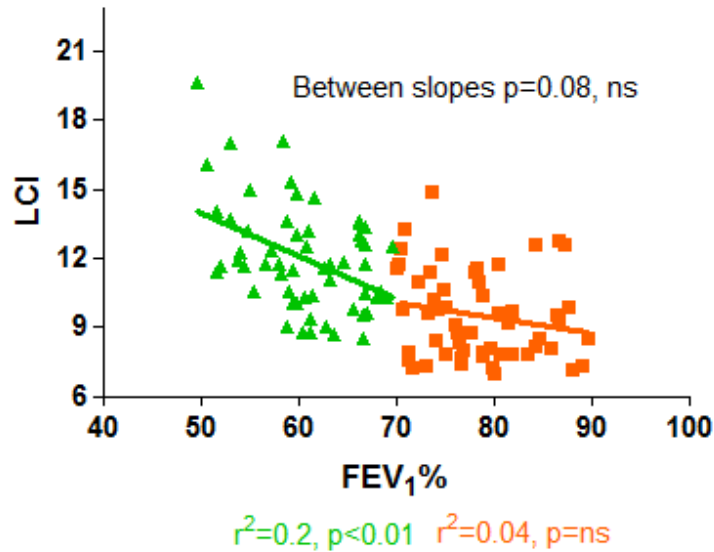


Figure 5.6 Graph of FEV₁% and LCI from the whole group at baseline, split by severity. The more severe group (<70% FEV₁ in green) have a significant regression line ($p<0.01$) but in the milder group (>70% FEV₁ in orange) the regression line between FEV₁% and LCI in both groups is not significant ($p=ns$).

5.8.2 Pre dosing correlations

One possible explanation considered for the lack of treatment effect in LCI was that there were differences between the two groups at baseline. Mean ($\pm$ SD) pre LCI in the placebo group (52 subjects (unable to obtain a pre LCI result for 2 patients due to trace exclusion criteria)) was 11.2 ± 2.5 . LCI in the active group was 10.5 ± 2.4 (61 subjects (unable to obtain a pre LCI result for 1 patient due to trace exclusion criteria)). LCI was not different between groups ($p=0.2$). As expected, LCI values were well correlated to FEV₁% (Figure 5.7), in both groups ($p<0.0001$). Regression slopes for placebo and active groups were not different. All LCI values were abnormal, none of them being below the HC upper limit of normal (ULN) for LCI (ULN was 7.5 as per the HC LCI data from the SF₆ modified Innocor results in Chapter 4). When splitting the groups by severity (<70% or >70% FEV₁) there were no differences between active and placebo groups in pre dosing LCI (<70% 11.76 ± 2.4 in active vs. 12.22 ± 2.18 in placebo, $p=ns$. >70% 9.2 ± 1.6 in active vs. 9.92 ± 2.2 in placebo, $p=ns$). Linear regression slopes for both groups remained significant in those with worse lung function and were not different from each other. In the milder patients the slopes between LCI and FEV₁% were not significant in either treatment group.

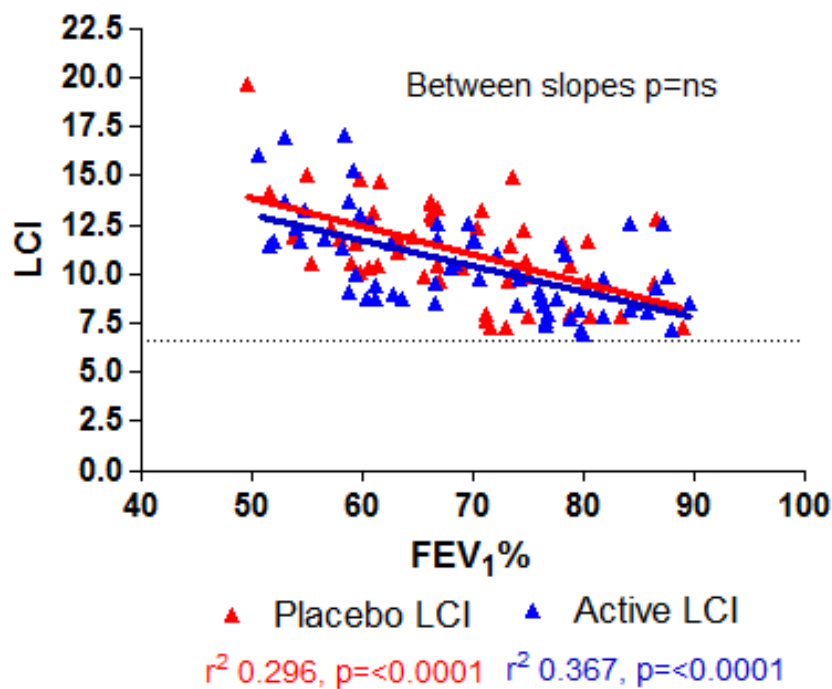


Figure 5.7 Pre-dosing correlations of FEV₁% and LCI in both active and placebo groups. As expected, the lower the FEV₁% the higher the LCI. The dotted line represents the upper limit of normal for healthy control patients; no CF patients in the MDT fall within the normal LCI range. Regression slopes between groups are not significantly different.

Similar pre dosing correlations were performed for 4 outcome measures (MEF₂₅₋₇₅, CT small mucus plugs, large mucus plugs and gas trapping) against FEV₁% and LCI. These other outcome measures were selected on the basis of reflecting small and large airway disease. All of the regression lines were significant (Figure 5.8). For the whole group of subjects, LCI and FEV₁% correlated to a similar extent with all of the outcome measures. As multiple comparisons were completed and the main trial results were not significant for a number of these secondary outcomes, I only regarded a p value of <0.01 as significant to ensure appropriate power between groups. There were no significant differences between active and placebo groups when MEF₂₅₋₇₅ or CT parameters were correlated with FEV₁% or LCI (Figure 5.8).

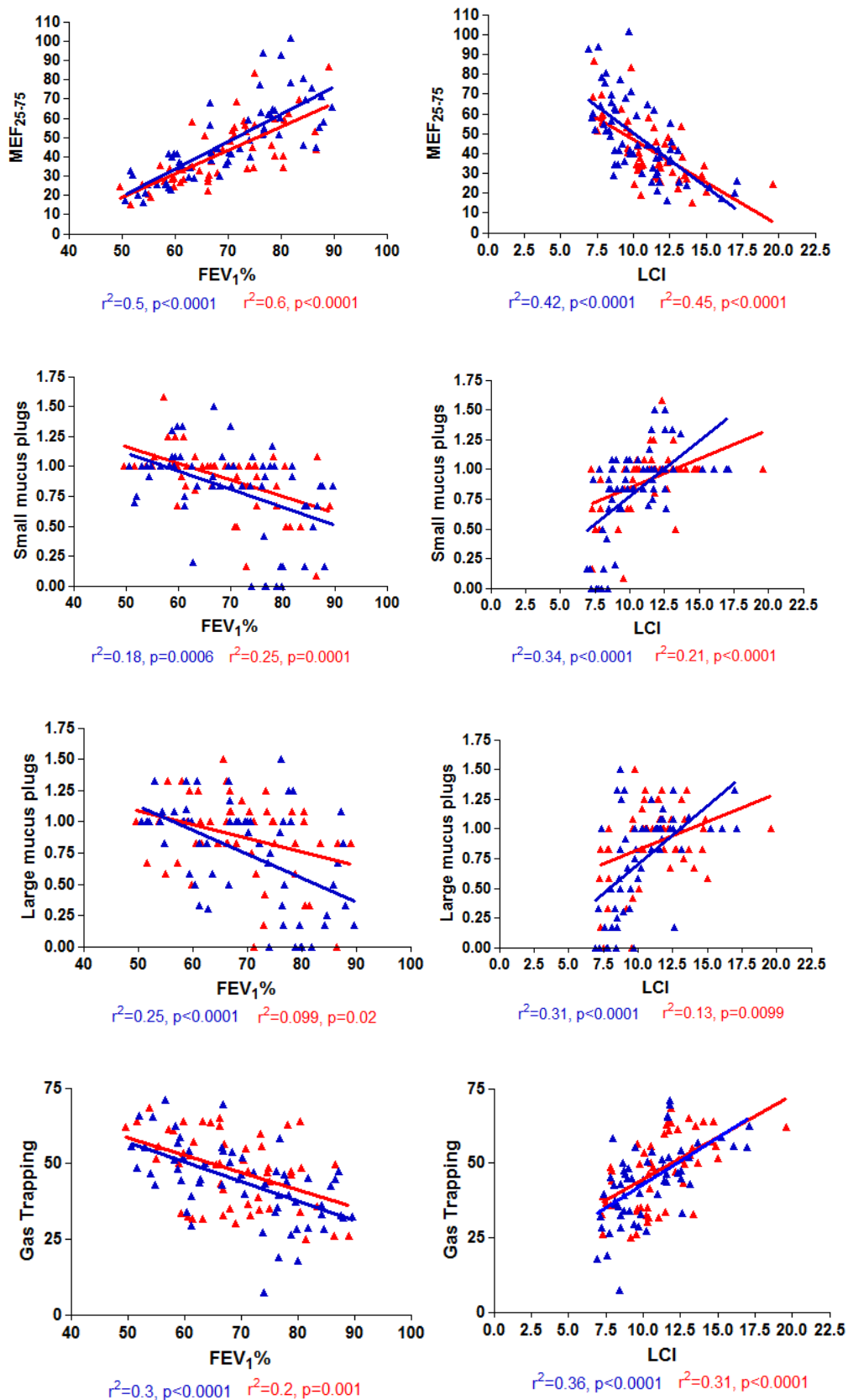


Figure 5.8 Pre dosing correlations of FEV₁% and LCI with other outcome measures, MEF₂₅₋₇₅, small mucus plugs, large mucus plugs and gas trapping. All of the regression slopes are significant; there are no differences between groups. For CT parameters the r^2 is higher for LCI than FEV₁%.

5.8.3 LCI change over the trial

Having shown that the two groups were similar at the start of the trial in their LCI values and the way these related to other measures, I next went on to explore further changes in LCI over the course of the trial which were shown in brief in section 5.4.3. LCI results were explored a) by severity, b) at an earlier time point (Dose 4) including LCI coefficient of variation across the trial to determine whether the within test variability contributed to any differences seen between groups c) by plotting the individual delta change across the trial.

5.8.3.1 LCI and severity

In the same severity subgroups as FEV₁% i.e. 50-70% and 70-90%, LCI was plotted against treatment visit (Figure 5.9). In both severity groups LCI deteriorates over time but in the active group (both severities) there is possible delayed deterioration compared to placebo.

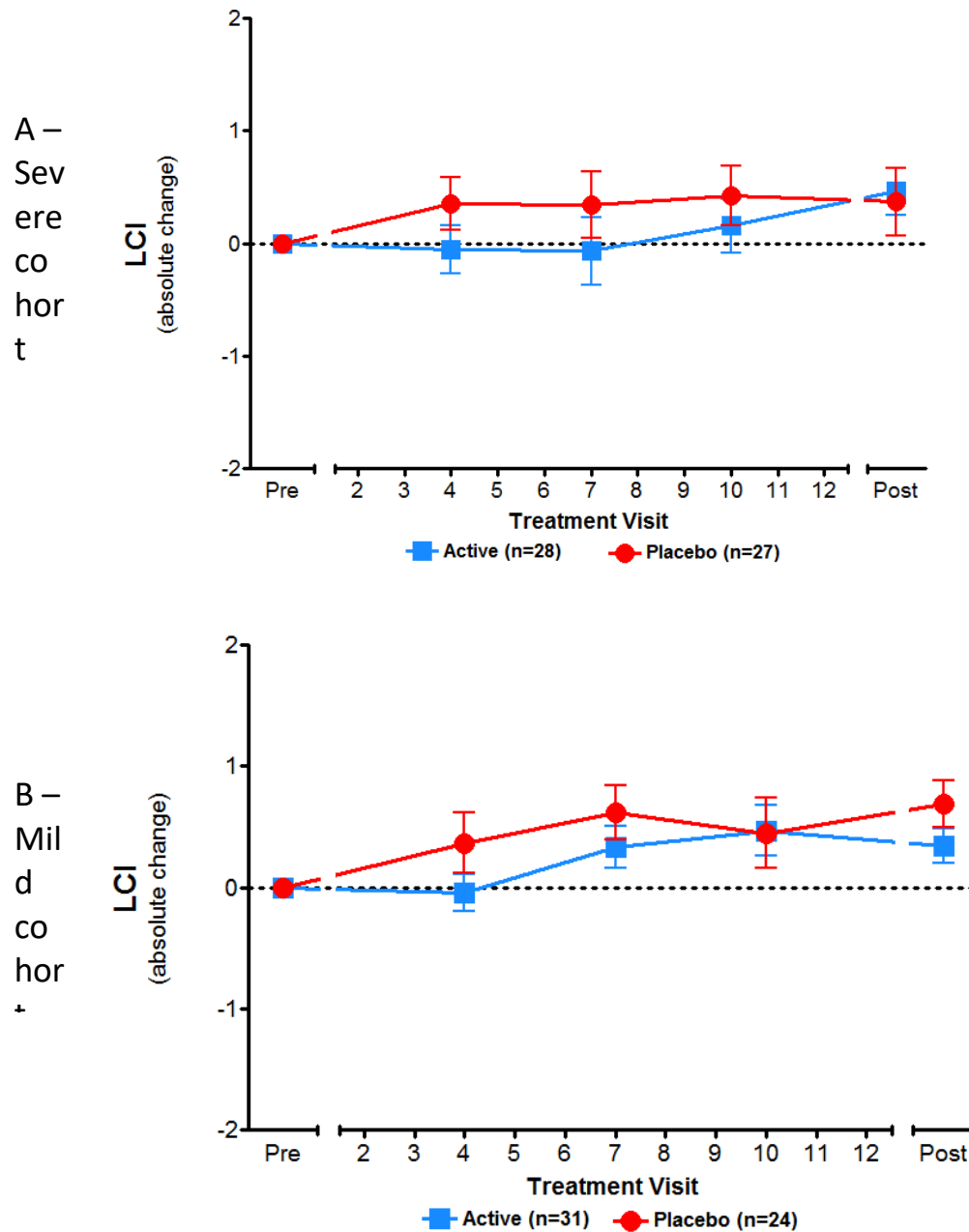


Figure 5.9 Time course of the LCI response stratified by baseline FEV₁% at trial entry. Error bars indicate standard error of the mean. A. Baseline FEV₁% between 40-70% predicted. B. Baseline FEV₁% between 70-90% predicted. A. The treatment effect was 0.05 (95% CI -0.7-0.8), B. The treatment effect was -0.4 (-0.9-0.04). There was no difference between groups $p=0.3$.

5.8.3.2 LCI results at an earlier time point

LCI progression over the whole trial did not show a statistically significant difference between groups; this can be seen in Figure 5.5 where data points post trial are similar and have overlapping error bars. However, we observed that at dose 4, the lines diverged, active

treatment group remaining stable whereas those from the placebo group worsening (also shown in the severity analysis above). This led me to perform a post-hoc, exploratory analysis of this early 3-month time period to determine if there was an early LCI signal that was lost over the course of the trial and any reasons why this might be the case. As previously described, there were no differences in LCI between active and placebo groups at baseline (demographics section). However, LCI was significantly different between active and placebo groups at DO4 ($p=0.01$) (Figure 5.10). This was not the case for FEV₁% (Figure 5.10).

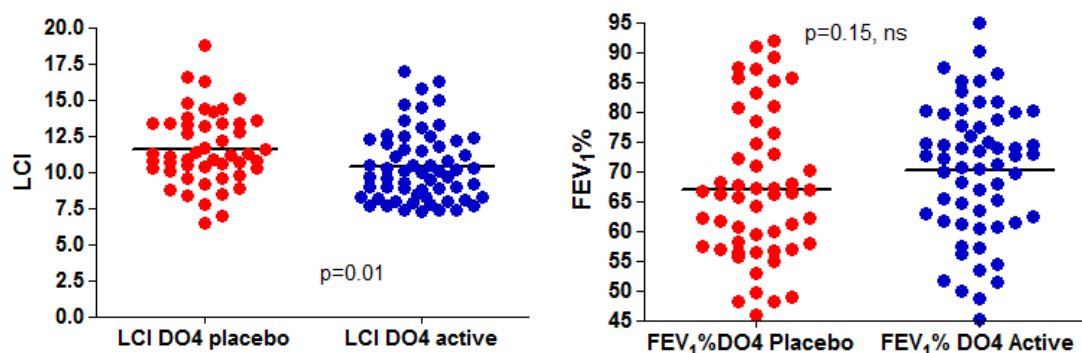


Figure 5.10 Comparison of LCI and FEV₁% values for Active and Placebo groups at Dose 4.

Comparison of the delta changes in FEV₁% and LCI at this early time point show that the two physiological outcome measures are not correlated (Figure 5.11). I also looked at whether the early appearance of an apparent LCI response to treatment correlated with the treatment effect seen in FEV₁% at the end of trial (correlation of LCI delta change from pre dosing to dose 4 vs. FEV₁% delta change from pre dosing to FU). There were no correlations in either group between early LCI responses vs. FEV₁% trial responses (placebo $p=0.8$, ns, active $p=0.4$, ns). This is an interesting finding but may imply that this exploratory data is underpowered and does not translate to meaningful information.

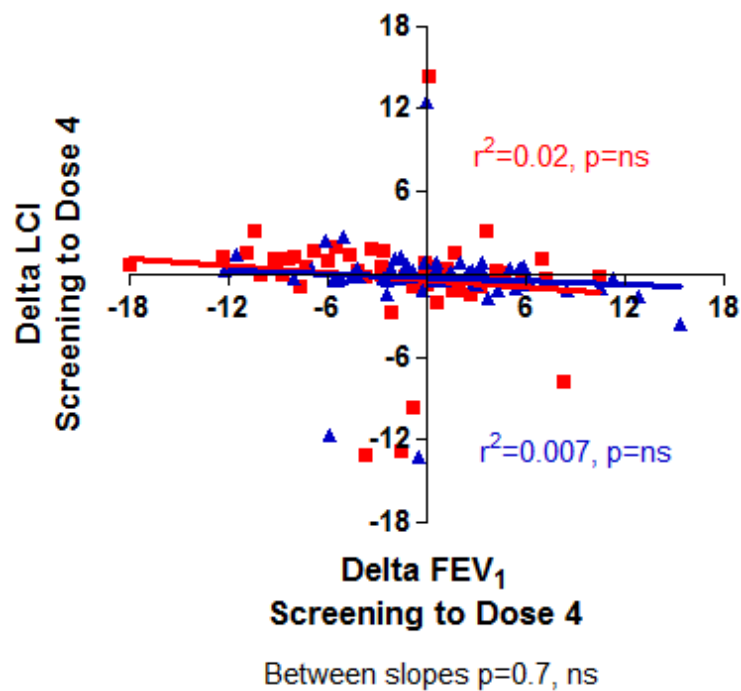


Figure 5.11 Linear regression analysis of delta changes in FEV₁% and LCI from pre dosing to dose 4 visits in both treatment groups. The regression lines are not significant showing that delta changes in FEV₁% and delta changes in LCI are not related.

5.8.3.2.1 Was the difference at Dose 4 more pronounced in one severity group?

An interesting pattern was observed when the group was split by baseline severity (<70% and >70% FEV₁): the difference at Dose 4 was observed in the milder patients only (9.19 ± 1.5 in active vs. 10.27 ± 2.2 in placebo, $p=0.04$), not in those with worse disease (11.74 ± 2.6 in active vs. 12.62 ± 2.3 in placebo, $p=0.4$, ns) (Figure 5.12). For the whole population, the delta change in FEV₁% from pre dosing to dose 4 between groups was also significantly different (Active: 0.3 ± 5.5 vs. Placebo: -1.9 ± 5.8 , $p=0.04$).

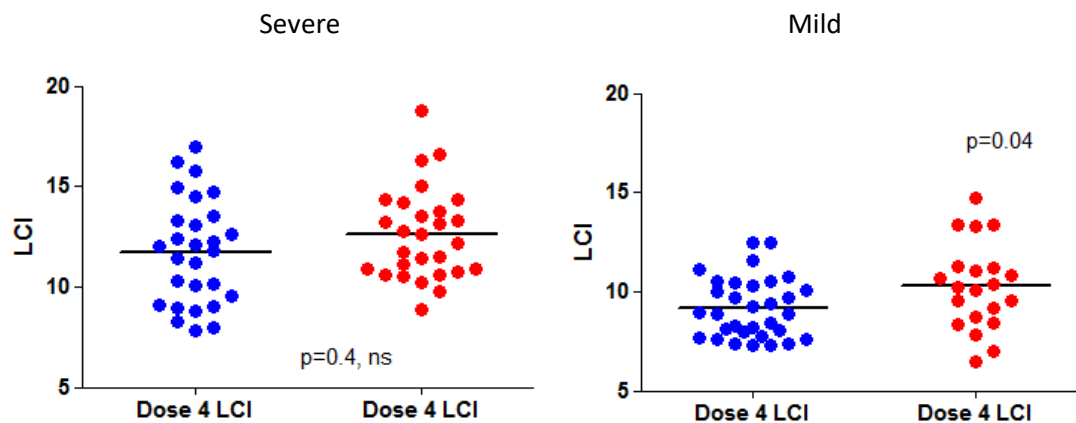


Figure 5.12 Comparison of Active and Placebo dose 4 LCI results in those patients with severe ($<70\%$ FEV₁ on the left) or mild disease ($>70\%$ FEV₁ on the right). There were no differences in LCI pre dosing but at dose 4 milder patients had different LCI results between groups.

5.8.3.2.2 Was the difference due to stability at the beginning of the trial?

As LCI has previously been shown to be more sensitive than spirometry in mild-moderate disease, I wondered whether this was the explanation for these early changes and that LCI was picking up a treatment response at DO4, which was particularly marked in the milder subgroup. This raised the question of why it didn't also show the effects later in the trial. Stanojevic *et al.* (2016) found that the LCI interclass correlation coefficient, that compares the variance between subjects with repeated measures, increased when preschool patients were considered to be stable therefore implying that some of the within test variability of LCI depends on patient symptoms and clinical status [259]. One possibility for this relates to stability at the various time points, which I next attempted to explore.

5.8.3.2.2.1 Within test variability

Within test LCI CV% was low for both groups and not statistically different between groups at any visit ($p>0.01$) (Figure 5.13). When exploring if LCI CV% was affected by severity there were no significant linear regression lines therefore it is difficult to make assumptions based on this exploratory data.

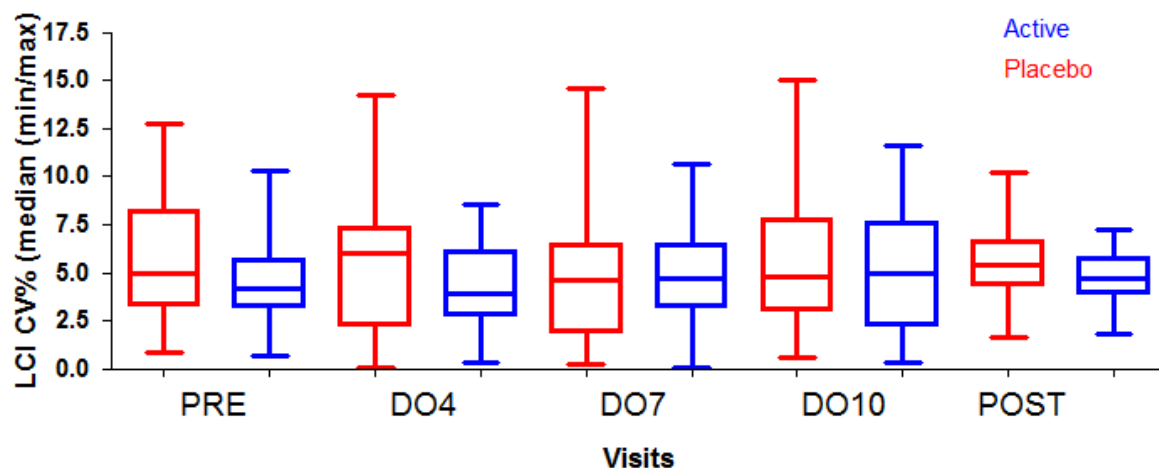


Figure 5.13 Within test LCI coefficient of variation (CV%) for Active and Placebo groups at each visit. A student t test revealed that LCI CV% was significantly different (higher for placebo group) pre dosing, but between groups LCI CV% was not different for the rest of the trial. ANOVA analysis of LCI CV% of each group across the trial was not significantly different.

5.8.3.2.2 Between test variability

Between test LCI CV% before and after dose 4 were used to assess the stability of the patients at this earlier time point. Two approaches were used a) calculate between test LCI CV% from all available visits before and after dose 4 (despite there being more visits after dose 4), b) calculate between test LCI CV% before and after dose 4 from the same number of visits. Non parametric Wilcoxon matched pairs test show FEV₁% CV% are significantly lower pre DO4 compared to after DO4 when all available visits are utilised ($p < 0.001$), LCI CV% is nearing significance ($p = 0.06$) with the same visit number (Table 5.7). This perceived variability is no longer apparent when the same number of visits before and after DO4 is assessed; highlighting that variability is increased with increased test number. These data are difficult to interpret since the utility of LCI CV% as a marker of stability is not fully understood but it does not seem to influence the earlier LCI response to treatment.

Table 5.7 LCI and FEV₁% coefficient of variation results pre and post Dose 4. The first two rows show data when incorporating all of the available visit data pre and post Dose 4. The last two rows show data from the same number of visits to ensure the variability is not to with increased test numbers.

	Placebo	p value	Active	p value
LCI CV% Pre dosing visits to Dose 4 vs. Dose 7 to follow up (median (min-max)).	3.67 (0.04-17.69) vs. 4.97 (0.04-14.64),	0.06	2.37 (0.14-11.55) vs. 4.17 (0.48-13.6)	<0.001
FEV₁% CV% Pre dosing visits to Dose 4 vs. Dose 5 to follow up	3.83(0.54-11.96) vs. 5.31(2.31-9.85)	<0.01	4.31(0.8-15.6) vs. 4.7(2.14-16.83).	<0.01
LCI CV% SCR and DO4 vs. DO10 and FU (same visit number pre and post dose 4).	3.7 (0.05-17.7) vs. 4.1 (0.02-18.02)	ns	2.37 (0.1-11.6) vs. 2.7 (0.2-14.6)	ns
FEV₁% CV% SCR, DO2, DO3, DO4 vs. DO10, DO11, DO12, FU (same visit number pre and post dose 4).	3.8 (0.5-11.96) vs. 3.96 (0.9-14.3))	ns	4.3 (0.8-15.6) vs. 3.8 (0.6-21.3)	0.05

5.8.3.2.3 Was there a difference in LCI at an earlier time point because patients had not yet experienced a pulmonary exacerbation?

The median (range) time to first exacerbation was 100 days in placebo (range 11 to 332, 7 people did not have an exacerbation but were still included in the median calculation) and 86 days in active (range 6 to 331, 6 people did not have an exacerbation but were still included in the median calculation), there was no difference between groups (p=ns). Dose 4 occurred after approximately 112 days (28 days from screening and between dosing visits x 4) therefore approximately half of the individuals had had an exacerbation by Dose 4 and there were no differences between active and placebo groups (Figure 5.14). Half of the

patients had experienced an exacerbation before DO4 so this does not explain the earlier LCI signal.

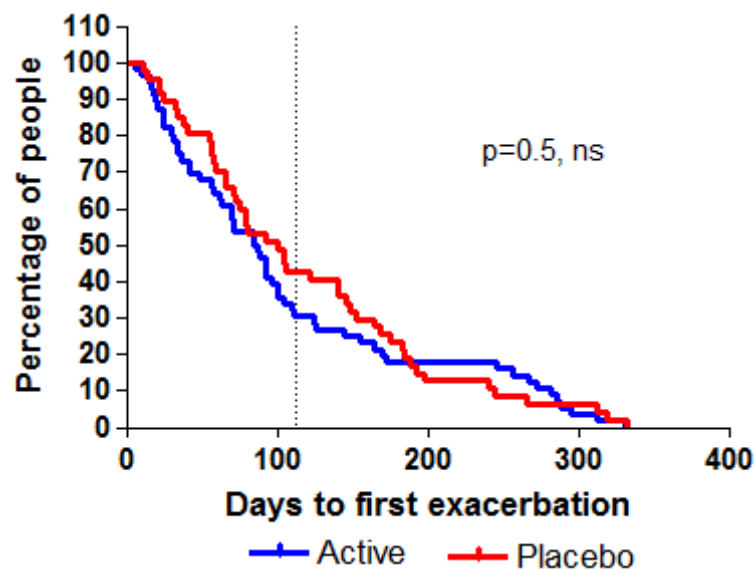


Figure 5.14 The proportion of people and the number of days until their first exacerbation.

The dashed line represents 112 days (approximating to the DO4 visit). There was not a statistically significant difference between active and placebo groups.

Significant LCI results at an earlier time point during the trial do not seem to be explained by between test variability and stability in terms of people not experiencing their first exacerbation during the trial.

5.8.3.3 Delta changes in LCI

Next, LCI data was organised according to the absolute delta change of LCI across the trial (in line with FEV₁% data in the trial paper, Figure 5.9). 33.8% of active patients' LCI improved (decreased) by a median (range) of 0.59 (0.04 to 1.94) LCI points. 66.1% of patients had worse LCI by the end of the trial (-0.9 (-2.66 to -0.02)) (Figure 5.15).

29.4% of placebo patients' LCI improved (decreased) by a median (range) of 0.96 (0.03-3.305) LCI points. 70.6% of patients had worse LCI over the trial (-1.14 (-3.31 to -0.02)) (Figure 5.9). There were no differences between groups.

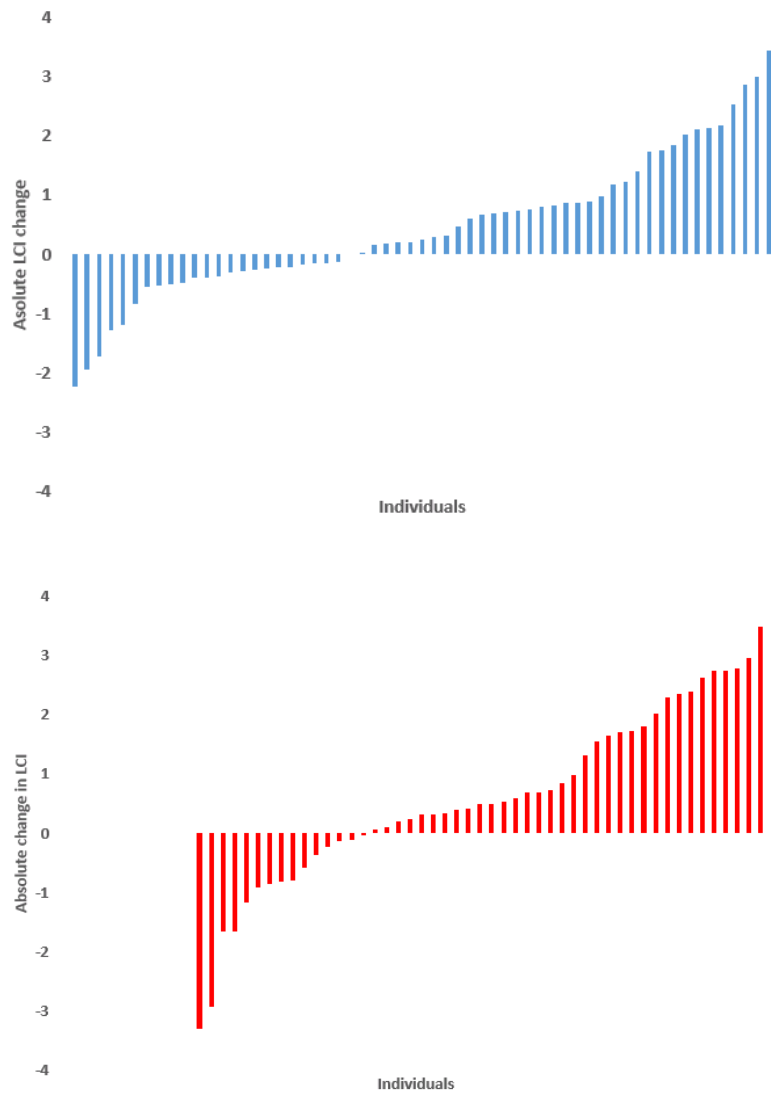


Figure 5.15 Active (blue) and Placebo (red) LCI waterfall plot showing the absolute change of LCI over the trial for all individuals. Negative delta means the LCI improved, positive delta means the LCI got worse. The majority of patients in both groups deteriorated in this outcome.

5.8.3.3.1 How did changes in LCI and FEV₁ relate?

FEV₁% showed a treatment effect in the trial but LCI did not. I wished to explore possible reasons for this, for example the large airways improving more than the small airways.

5.8.3.3.1.1 Post Trial Correlations

I first performed correlations between post dosing FEV₁% and LCI data from active and placebo patients using linear regression analysis (Figure 5.16) to assess whether the

relationship between FEV₁% and LCI was still maintained. The same was then completed between LCI and the other 4 outcome measures.

At the end of the trial, regression slopes between FEV₁% and LCI were statistically significant for both active and placebo groups but were now significantly different from each other (Figure 5.16 p=0.02). In both groups the slope was similar to that obtained at pre-dosing, but the movement between groups seemed to come from progression of lung disease in the placebo group where the post dose correlation between FEV₁% and LCI was less significant than at pre dosing. Due to the exploratory nature of this analysis and lack of power for LCI results, this is difficult to interpret.

All of the post dose regression lines were significant when comparing LCI and FEV₁% against the 4 other major secondary outcome measures (Table 5.8). They had the same directional magnitude as pre dosing (worse with high LCI or lower FEV₁%). There were no significant differences between active and placebo groups in any post-trial correlation with other outcome measures.

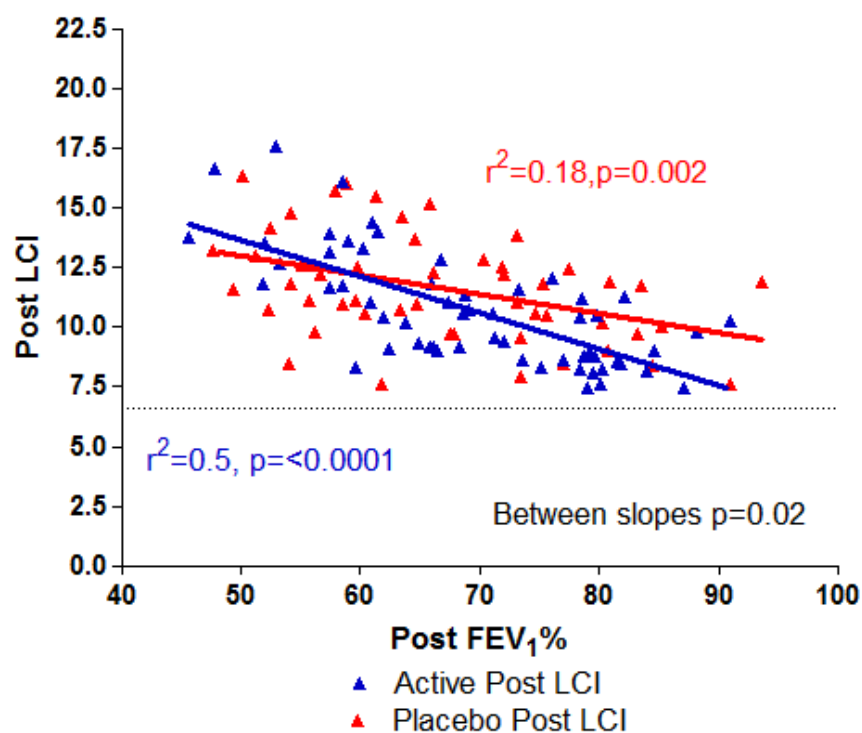


Figure 5.16 Post-dosing correlation of FEV₁% and LCI in both active and placebo groups. No patients fall within the normal LCI range (under the dotted HC ULN line). Regression slopes between groups are significantly different (p=0.02).

Table 5.8 Post trial linear regression slopes for FEV₁% and LCI vs. other secondary outcome measures (MEF₂₅₋₇₅ and CT parameters, large mucus plugs, small mucus plugs and gas trapping). All of the regression slopes were significant. There were no differences between active and placebo groups.

	FEV ₁ %	LCI
MEF₂₅₋₇₅		
Active	$r^2=0.06, p<0.001$	$r^2=0.47, p<0.001$
Placebo	$r^2=0.7, p<0.001$	$r^2=0.31, p<0.001$
Large Mucus Plugs		
Active	$r^2=0.3, p<0.001$	$r^2=0.3, p<0.001$
Placebo	$r^2=0.1, p<0.01$	$r^2=0.1, p=0.01$
Small Mucus Plugs		
Active	$r^2=0.2, p<0.001$	$r^2=0.3, p<0.001$
Placebo	$r^2=0.2, p=0.001$	$r^2=0.2, p<0.01$
Gas Trapping		
Active	$r^2=0.4, p<0.001$	$r^2=0.4, p<0.001$
Placebo	$r^2=0.2, p<0.001$	$r^2=0.3, p<0.001$

5.8.3.3.1.2 Did FEV₁% and LCI change similarly across the trial?

Delta changes of FEV₁% and LCI were compared in individuals to see if FEV₁% and LCI changed in the same direction during the trial. Had they done so, this might suggest that LCI was less precise than FEV₁ and underpowered; if FEV₁% and LCI were not aligned then this may show that LCI was not impacted by gene therapy treatment. For ease here, a negative change represents deterioration both for FEV₁% (FEV₁% decrease) and LCI (LCI increase). Delta LCI and FEV₁% correlated significantly for the placebo group ($p<0.0001$ – those with negative changes in FEV₁% also had negative changes in LCI) but in the active group there was no correlation (Figure 5.17), underscoring the observation of a significant treatment effect on FEV₁, which was not seen in LCI.

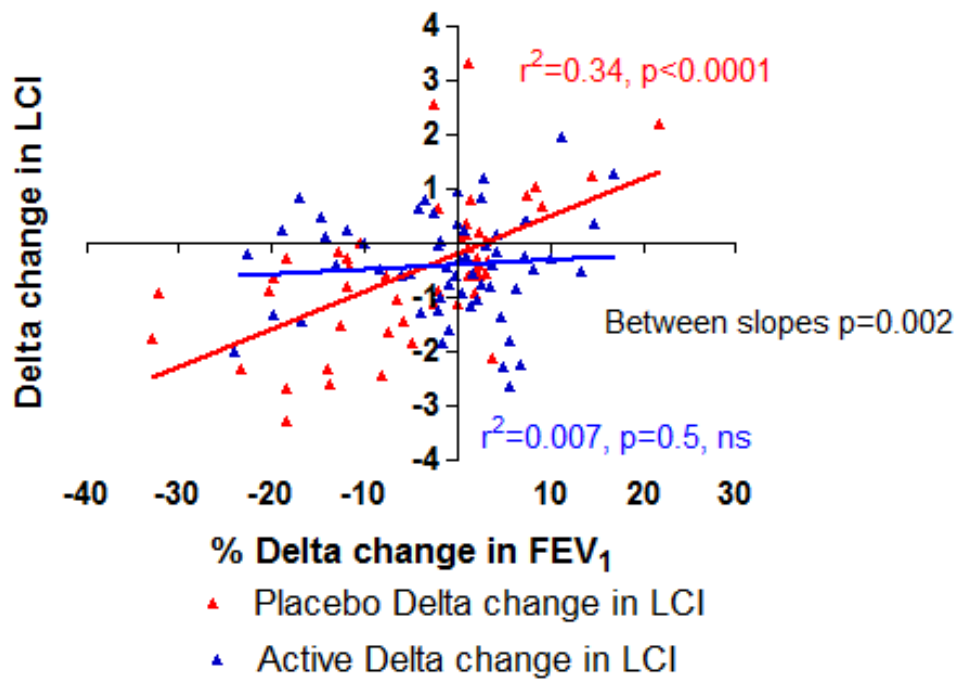


Figure 5.17 Percentage delta changes in FEV₁% vs. absolute delta changes in LCI. The placebo group had a significant regression line ($p<0.0001$) but in the active group there was no correlation.

5.8.3.3.1.3 Did a subject's starting lung function affect their delta change across the trial?

To see whether baseline severity impacted the degree of change available after treatment, baseline lung function and delta change were plotted together. Despite the appearance (albeit, not reaching statistical significance) of a greater FEV₁% response in the group of patients with lower lung function (50-70%), on an individual basis, baseline FEV₁% did not relate to FEV₁% change across the trial (Figure 5.18). The regression lines for active and placebo are not significant, nor are they different from each other.

The same graph was completed but using starting LCI value. Again, for ease of comparison, a negative change represents deterioration for LCI (LCI increase). In the placebo group, the lower the LCI at the beginning of the trial the more likely there was to be a deterioration change in LCI (increased LCI) (Figure 5.19); the same relationship was not seen in the active group, although the data plot was noisy and comparison of the regression lines did not reach statistical significance, so this finding should be interpreted with caution.

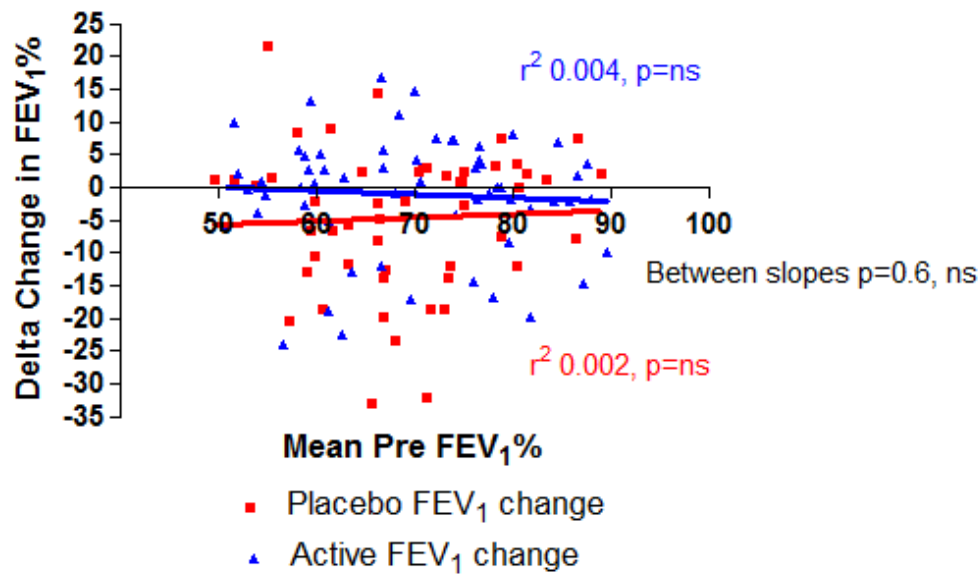


Figure 5.18 Starting FEV₁% vs. FEV₁% delta change. Regression lines were not significant for either group and were not different to each other.

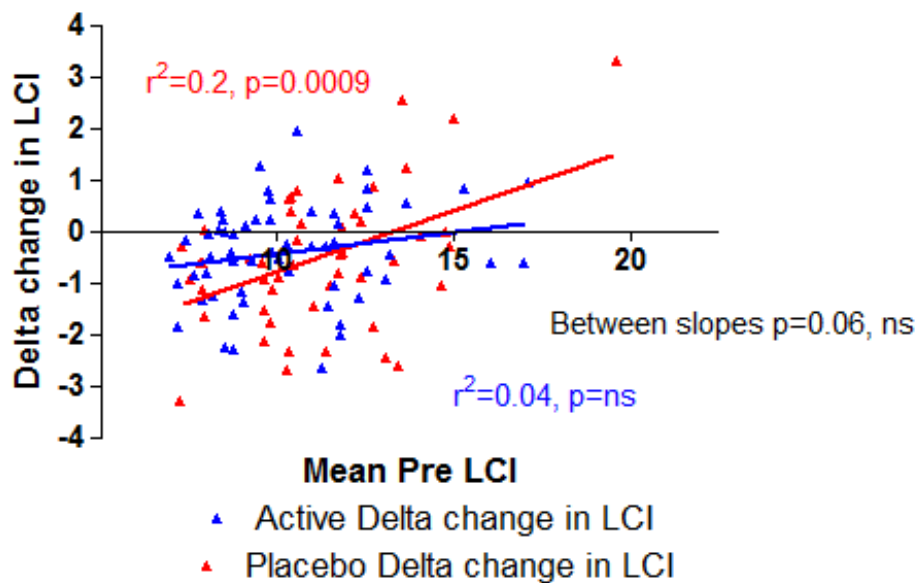


Figure 5.19 Starting LCI vs. LCI delta change. The regression line for placebo was significant ($p = 0.0009$) but very weak; that was not the case in the active group ($p = ns$). In the placebo group, the lower the LCI the more deterioration in LCI (increased LCI), the higher the LCI the more improvement (lower LCI). The linear regression lines between groups were not significantly different.

To quantify these patterns of change and explore whether LCI and FEV₁% changed to the same magnitude in both treatment arms, delta changes of FEV₁% and LCI across the trial were ranked from the most improved to the most deteriorated. Each individual subject's FEV₁% and LCI delta change was then linked and movement between quartiles were recorded. Figure 5.20 shows that most subjects had delta changes in FEV₁% and LCI that were similar and kept the individual in the same quartile (zero column). Active treatment seemed to promote more movement between quartiles both in a negative and positive direction of change (chi squared of percentage of people vs. movement between quartiles of ranked delta change, $p=0.0003$). This once again illustrates that for the active group there was disassociation between FEV₁% and LCI.

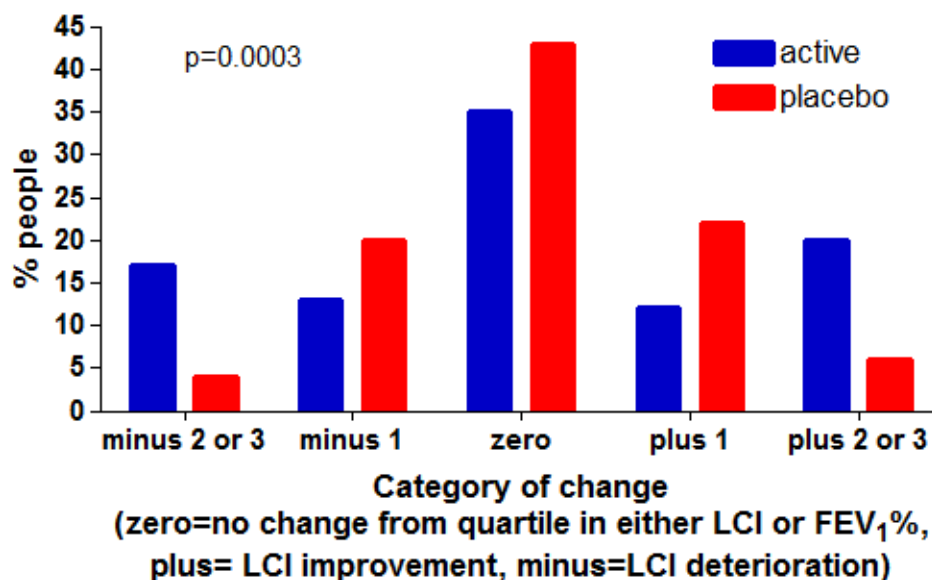


Figure 5.20 Comparison of both delta changes in FEV₁% and LCI. Delta changes of FEV₁% and LCI across the trial were ranked from the most improved to the most deterioration. Each individual subject was then linked and movement between quartiles was recorded. The y-axis is the percentage of people in each treatment arm and the x-axis is the movement from each quartile. Chi squared analysis showed significant differences between groups.

5.8.4 Correlation with other outcomes

As the relationship between FEV₁ and LCI changed in the trial in the placebo group, it may be that the other small and large airways readouts would also no longer correlate, either due to the change of FEV₁% seen post trial and or the lack of LCI response to treatment.

5.8.4.1 *Physiological outcome measures*

Post-trial MEF₂₅₋₇₅ and CT parameters (small mucus plugs, large mucus plugs and gas trapping) remained significantly correlated to both FEV₁% and LCI (pre trial results in Figure 5.8). This may indicate that the treatment was not effective enough to show a change in these outcome measures and therefore they remained in line with FEV₁% and LCI. It could also show that the outcome measures are not sensitive enough to any treatment changes. To explore this further the delta changes in these outcome measures were correlated with FEV₁% and LCI.

Changes in FEV₁% and LCI were most closely represented by MEF₂₅₋₇₅. Linear regression analysis of small mucus plugs with FEV₁% and LCI were also significant for the active group. There were no differences in large mucus plugs and gas trapping (Figure 5.21).

Further analysis revealed that delta changes in FEV₁% were most closely matched by MEF₂₅₋₇₅ (unsurprisingly as they are derived from the same test), then LCI, then to similar extents the CT markers. When tracking/matching delta changes in these outcomes with the FEV₁% change (similar to Figure 5.20), only four patients in the active group and one patient in the placebo group had similar delta changes in all outcome measures (data not shown).

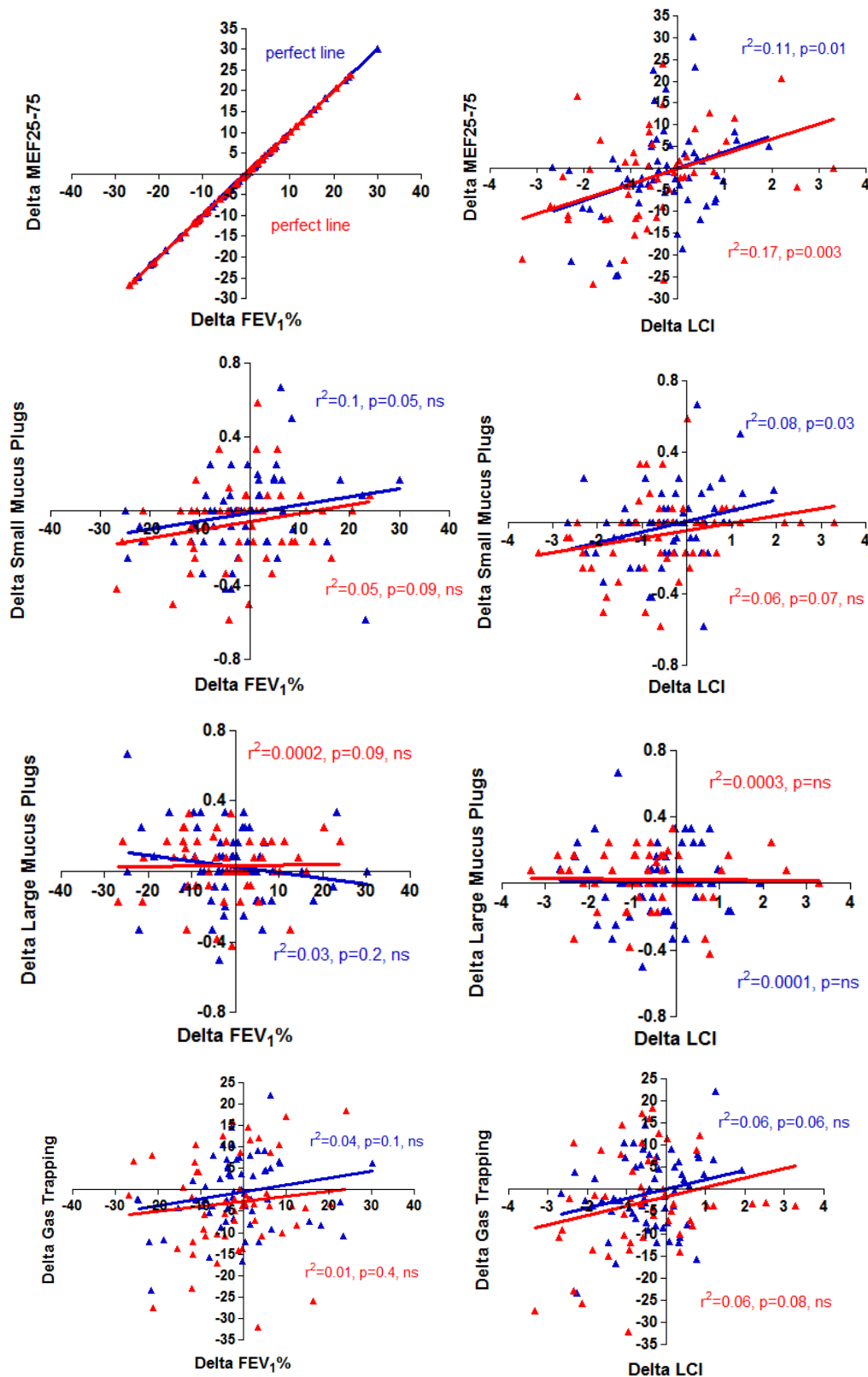


Figure 5.21 Linear regression analysis of delta FEV₁% and LCI with MEF₂₅₋₇₅, small mucus plugs, large mucus plugs and gas trapping. Only MEF₂₅₋₇₅ was significantly related to FEV₁% and LCI. There were no differences between active and placebo groups in any outcome measure.

5.8.4.2 CFQR QoL

An MCID of 5.6 points has been established in CFQR QoL [81]. Although there was no treatment effect observed in CFQR QoL, I wished to use changes in this parameter as a means of benchmarking change in FEV₁ and LCI. This section seeks to determine if those people who improved above the MCID for CFQR QoL had similar changes in FEV₁% and LCI. If so, then the results could be helpful in defining MCIDs for these physiological tests.

Across the trial, the respiratory domain from the CFQR did not change significantly between active and placebo groups. The placebo group did have a higher respiratory score pre dosing (81.89 ± 12.68 vs. 76.56 ± 14.55 , $p=0.04$) but this was not the case post trial (77.06 ± 15.46 vs. 76.65 ± 16.42 , $p=ns$). There was a trend for those with the best lung function having the best quality of life scores but correlations (assessed using $p<0.01$) between either FEV₁% or LCI with the CFQR respiratory domain were not statistically significant pre dosing. There were also no post-dosing correlations between CFQR respiratory and FEV₁% or LCI in either group.

A minimal clinically important difference in the respiratory domain for the CFQR QoL has been identified [81]. It represents a change of 5.6 points when scaled to 100, which is approximately one change in category (always, often, sometimes and never) on the questionnaire. Of those patients who had a positive change in line with the MCID in QoL, the corresponding positive or negative change in FEV₁% and LCI was assessed to address whether the physiological markers respond in the same way to CFQR QoL.

44% of the active group had QoL delta above 5.6. This group was split equally into subjects experiencing positive (48%) and negative (48%) changes in FEV₁%. For LCI, 23% had positive changes (improvement), 76.9% had negative. Thus, an improvement in CFQR > MCID was not reflective of improvement in physiological change in the active group. This is further shown by the ROC curves in Figure 5.23 where the area under the curve is small for LCI and FEV₁% showing little sensitivity or specificity of FEV₁% to predict the CFQR respiratory domain change.

In the placebo group 46% of people had QoL scores above 5.6; 59% of them also had an improved FEV₁% after the trial (41% deterioration) but only 29% an improvement in LCI (71 deteriorated) (Figure 5.22).

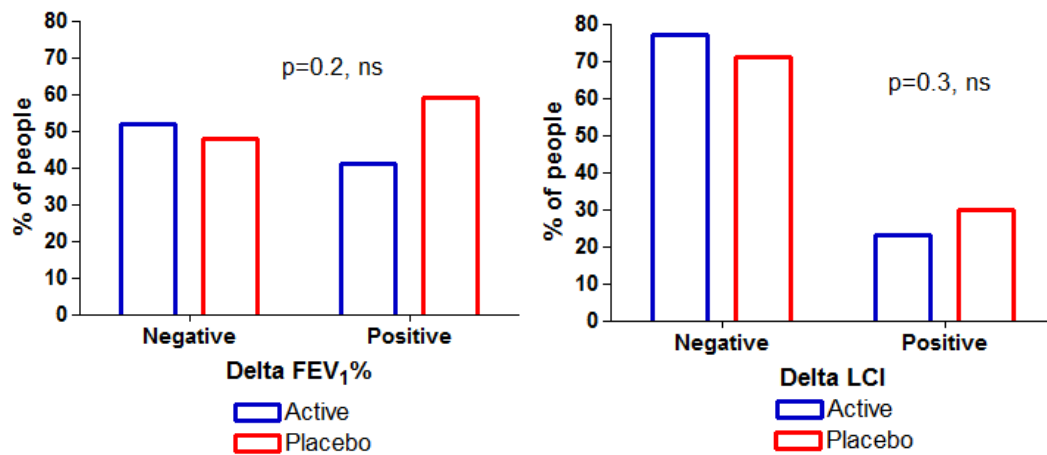
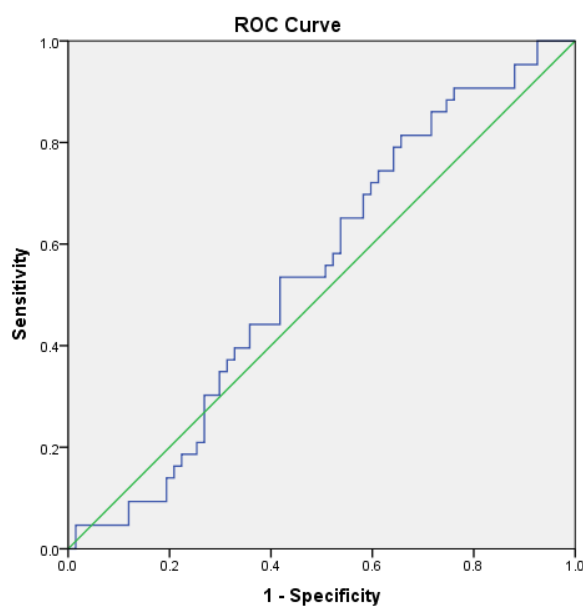


Figure 5.22 Those patients with quality of life scores above the minimal clinically important difference were selected. The percentage of people with improvements or deterioration in FEV₁% and LCI were then plotted to determine if physiological outcome measures were similar to patient reported outcomes. A Fisher's exact test was completed to determine the differences between groups and delta changes. More people had positive improvements in FEV₁% compared to LCI, but neither physiological outcome measure had statistically significant differences in the delta changes (p=ns) that were in line with CFQR QoL.

A



B

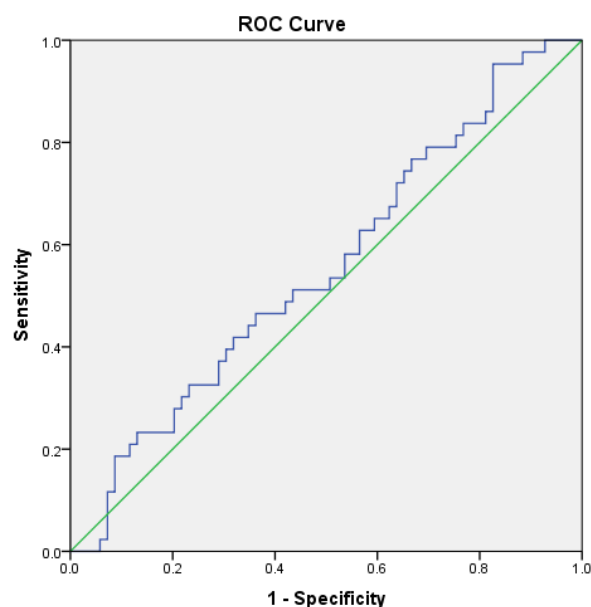


Figure 5.23 – Receiver Operating Characteristic (ROC) curves for data from both the active and placebo groups from the CFQR QoL Respiratory domain as a dichotomous variable (i.e. achieving the MCID or more (+5.6%) vs. the change in LCI over the trial (A) and change in FEV₁% over the trial (B). The area under the curve for LCI is 0.55 $p=0.4$, the area under the curve for FEV₁% is 0.56 $p=0.3$.

5.8.5 Run in results: did the MDT change the trajectory of FEV₁ and LCI?

As described, the Run-in study was started to find the optimal outcome measures for detection of gene therapy efficacy and safety. Longitudinal data were collected and a

proportion of the participants in the Run-in study went on to complete the MDT. For two years post completion of the trial, the GTC received local clinic letters for all of the participants. This enabled follow up data to be collected, including FEV₁%, although this was non-standardised and performed according to individual clinical procedures, and LCI was not collected at all. This section of the thesis aimed to explore whether active treatment in the MDT changed the trajectory of FEV₁% or LCI when compared to the placebo group. FEV₁% follow up data will also be explored for these patients. The group is small, so this is an exploratory analysis, and any findings from which would need interpreting with caution and confirming. 41 people from the run in study enrolled in the MDT, 7 of whom were excluded (reasons including the availability of ivacaftor and time commitments) leaving 34 patients (16 active treatment, 18 placebo) included in this analysis (demographic data in Table 5.9). There were no differences between demographic data for either active or placebo groups from the run in study to the MDT. Importantly, although the group was too small to expect a statistically significant effect, the groups behaved similarly during the trial period as did the whole trial cohort in showing a divergence of active and placebo subjects for FEV₁% (Figure 5.23, section 0-13 months).

Table 5.9 Demographic data from 34 people who were in the run in study and went on to complete >9 doses of gene therapy or placebo doses during the multi-dose trial.

	Placebo (18)	Active (16)
Age (years) Mean $\pm$ SD or Median (range)	<18years n=4 (22%)(15.25 $\pm$ 1.09) >18years n=16 (88%) (30 (18 to 64))	<18years, n=4 (25%)(16.25 $\pm$ 1.29) >18years, n=12 (75%) (29.5 (18 to 50))
Sex	Male n=11 (61%) Female n=7 (38%)	Male n=10 (63%) Female n=6 (38%)
Centre	Ed n=11 (61%) Lon n=7 (39%)	Ed n=7 (44%) Lon n=9 (56%)
Height (cm) Mean $\pm$ SD	166.2 (10.8)	167.5 (12.05)
Weight (kg) Mean $\pm$ SD	61.17 (11.78)	67.4 (16.8)
FEV ₁ % Mean $\pm$ SD	67.8 (8.3)	69.3 (12.1)

Mean FEV₁% and LCI for all patients were plotted against time in 6 monthly intervals (Run in (-66 to 0), MDT (0-13) and FU (14-38)). All of the visits were anchored to the MDT. During the Run in and FU periods, if visits fell within the same 6 month period then the results were averaged. Though each individual had a similar number of visits throughout the whole monitoring period, due to the long time periods and less structured return for the next visit during run in and follow up, each dot of the graph may represent a different number of people. Linear regression slopes were plotted and active and placebo groups were then compared (Figure 5.23 and 5.24).

FEV₁%: for this small group, the trial itself didn't alter the change over time in FEV₁%; indeed slopes appeared flat. The apparent decline after the trial in the placebo group is hard to interpret, but it is of interest that we did not observe a decline in the active group upon 'withdrawal' of treatment.

LCI: Again in this small group, the trial itself did not alter LCI change over time. LCI appeared

relatively stable throughout the run in and MDT time periods, if not higher during the MDT potentially due to lung disease progression over time. It is unfortunate that FU LCI data had not been captured.

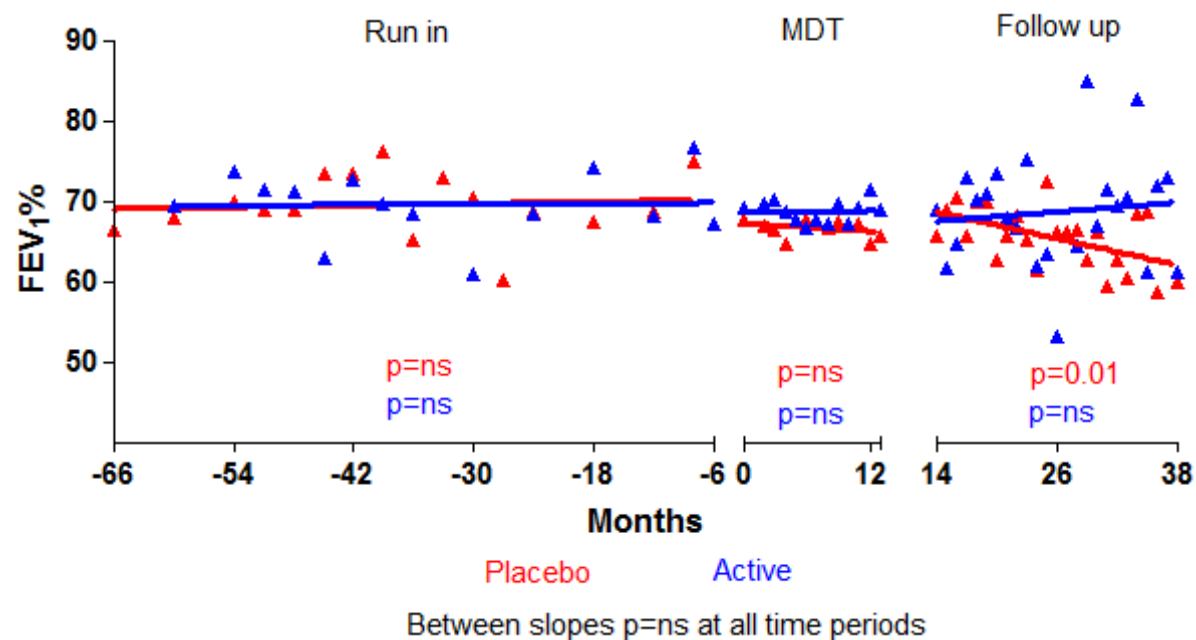


Figure 5.24 Graph to show the average FEV₁% from run in, multi-dose and follow up data in both active and placebo patients. The x-axis is split into 3 sections, representing the three different recording periods (-66 to -6 months is run in, 0-13 months is the multi-dose trial and 14-38 months is the 2 year follow up period). The linear regression lines were not significant for either group at any time period, apart from the placebo group during the follow up period. Active and placebo slopes were compared and were not significant at any time point.

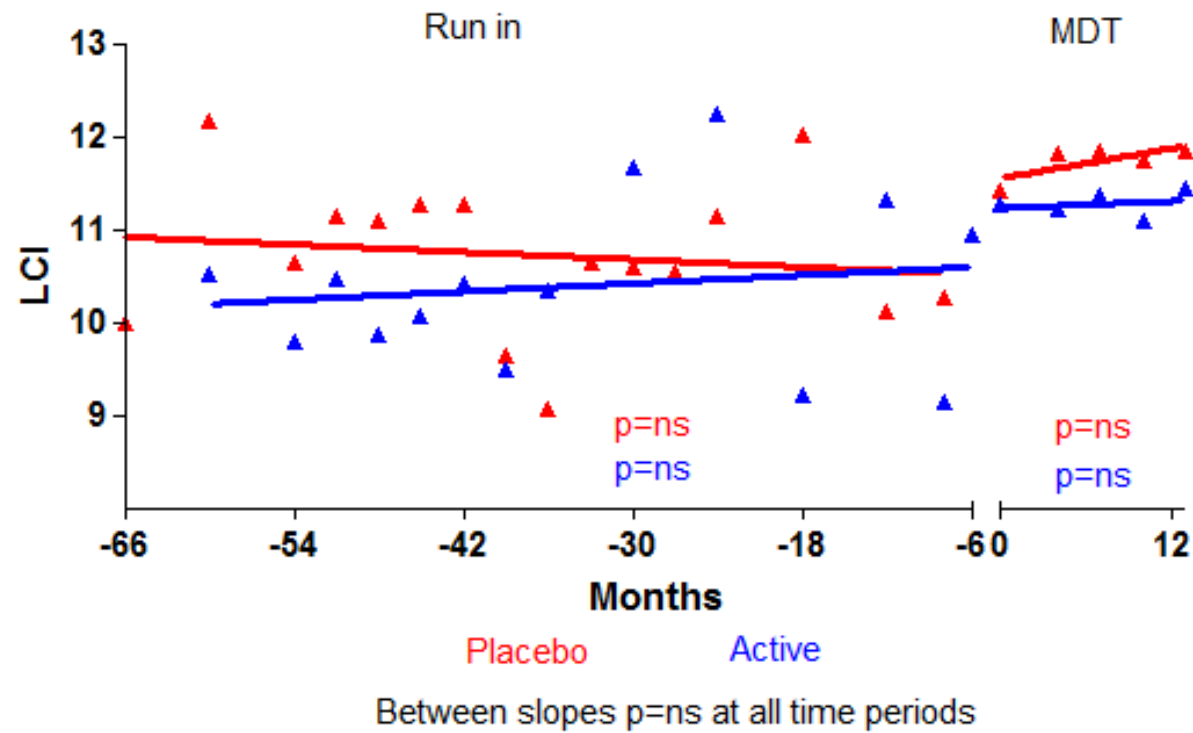


Figure 5.25 Average LCI from run in and multi-dose in both active and placebo patients. The x-axis is split into 2 sections, representing two different recording periods (-66 to -6 months is run in, 0-13 months is the multi-dose trial). The linear regression lines were not significant for either active or placebo groups. Active and placebo slopes were compared and were not significant.

5.9 Discussion

In this chapter, I have presented the main MDT outcomes as well as explored the outcome measure lung clearance index. LCI did not show a treatment effect post-trial and explanations for this have been evaluated. There are several areas worthy of further discussion, weaknesses, which will be considered, and opportunities for future work, which I will highlight.

5.9.1 Patient recruitment, retention and safety

The MDT is the largest gene therapy study of its kind, having progressed from an extensive cystic fibrosis gene therapy program [39]. The trial formulation was designed a) to permit repeated application over a period of time sufficient to determine change in clinically relevant outcomes, b) to minimise inflammatory responses by removing all CpG motifs and c) lead to extended duration of expression with a non-viral, humanised promoter. Overall, the drug was well tolerated and both groups experienced a similar number of adverse events. The adverse events that did occur during the trial were deemed to be of the nature expected in CF trials.

The MDT was a multicentre study powered for 120 patients, and the final PP dataset was 116 patients. A more detailed trial report including all aspects of the statistical analysis plan and full results produced by the consortium can be found here

<https://www.journalslibrary.nihr.ac.uk/eme/eme03050/#/full-report>.

16 patients from the active group and 8 patients from the placebo group discontinued the trial before completion. The reasons for withdrawal were variable from time commitments, travel issues, the availability of ivacaftor, development of exclusion criteria or withdrawing consent; the trial steering committee deemed these not to be study drug related. It highlights the issues faced when recruiting for large studies particularly in a competitive research environment where individual patients have the right to choose which studies to participate in and where large cohorts of patients are required to achieve the primary outcome [260]. The overall drop out rate seen here is similar to another long-term trial targeting the basic defect [44]. Opening of patient identification centres, incentives or shorter study durations for trials may help recruitment and retention but the small drop out only highlights the dedication and altruistic nature of the CF cohort.

5.9.2 Demographics

This thesis focuses on LCI, an outcome measure yielding expected values at baseline in this population. The mean pre-dosing $\text{Inn}/\text{SF}_6\text{LCI}$ of 10.78 ± 2.45 was above that seen from the HC population seen in Chapter 4 and other SF_6 HC (Table 4.1), so as expected LCI was able to identify disease.

In line with previous literature [141], $\text{FEV}_1\%$ and LCI were highly correlated. This graphical explanation of $\text{FEV}_1\%$ and LCI is not the first to show this relationship, Vermeulen *et al.* 2014 show a similar trend but their patients were younger and the negative correlation more pronounced [87]. This suggests the MDT patients were more severe, confirmed by no patients falling below the ULN LCI line in this study. It also shows the added benefit of LCI in the 16.4% of patients who had $\text{FEV}_1\%$ above 80% and yet LCI was already abnormal. The MDT trial enrolled children and some with preserved lung function; therefore it may be surprising that no subjects fell under the ULN for LCI. Healthy control reference ranges from a larger study population do not necessarily provide a larger range of values [105] but the difference between study group data only emphasizes the important requirement for robust LCI reference ranges so interpretation and clinical significance of a result is useful and precise (see section 3.5 in Chapter 3). In any case, interpretation of the MDT population at baseline using LCI ULN shows that by 12 years much of the lung damage that occurs with CF is already visible using LCI.

As well described in other groups, in this population LCI increases with age and severity of lung disease assessed by $\text{FEV}_1\%$, two factors which are clearly linked in this progressive disease. Interestingly, when splitting the patients by baseline severity, only those with worse lung function demonstrated a significant correlation in $\text{FEV}_1\%$ and LCI. In those with mild $\text{FEV}_1\%$ impairment, there was no correlation between LCI and $\text{FEV}_1\%$. This describes the added benefit of using LCI in these mild patients, as LCI is already abnormal despite normal spirometry values [73].

As would be hoped from well-matched, randomly allocated groups, there were no differences in these correlations between active and placebo groups. When LCI and $\text{FEV}_1\%$ were correlated with other small and large airways readouts the correlation was as expected and was of similar extents in both outcome measures. This shows that comparison of the active and placebo group post trial was valid.

5.9.3 Main trial outcomes (primary and secondary)

There was a significant ANCOVA adjusted treatment effect in the active group vs. placebo (Figure 5.3). The MDT met its primary endpoint, an FEV₁% treatment effect seen from 12 doses of pGM169/GL67A vs. placebo [181].

The mean difference in FEV₁% change across the trial was modest and was caused by deterioration in the placebo group and stabilisation (rather than improvement) in the actively treated group. There were no related variables to suggest specific phenotypes would be suitable for targeted treatment with this intervention except that those people with more severe FEV₁% at baseline seemed to improve more than those with milder disease (Figure 5.4). The GTC suggest that this increased treatment effect in more severe patients (although not statistically significant, but nor was the trial powered for this subgroup analysis) could be caused by increased deposition of pGM169/GL67A in the larger airways of those patients who had more peripheral obstruction.

Although not shown here, treatment effect analysis favored the active group (Forest Plots featured in the supplementary material [181]) and functional protein assessed by bronchial PD showed responses towards the normal range. However, evidence of transfected DNA was not apparent through mRNA measurement, suggesting that the selected dose may be too small to be quantifiable and further investigation increasing the dose should be explored.

This smaller response to treatment has also been seen in other clinical interventions [261, 262] and other studies targeting the basic CFTR defect [40, 44]. But these studies also had consistent improvements with treatment and supporting parameters including decreased pulmonary exacerbations or drug interactions explaining the primary outcome's limited response. They also targeted a specific population. The mantra of the GTC for patient selection "Can Deliver" (not too severe to prevent gene deposition) and "Can Measure" (not too well to mask clinical efficacy) [39] may well need to be revised to target apparent responders, change the cut off values or increase the dosing strategy. This may improve trial efficacy and consistency of response in any future trials. The pool of true placebo patients may also be reduced, particularly with the rapid clinical uptake of ivacaftor [263]. Although, there are only 4% of patients who could benefit from targeted ivacaftor therapy, the approval of Orkambi in F508del patients (even if it is not yet approved by the NHS) may increase the challenge of proficiently powered studies in the future. Future studies may have to include potentiator drugs in order to maintain the current clinical status of patients; but it may also be the case that these drugs have beneficial interaction treatment effects.

The magnitude and variability of the treatment seen in the MDT does not warrant immediate progression into phase III gene therapy trials [256] as further efficacy is required for example through increased dose or dosing frequency.

The difference between active and placebo groups may be more apparent because the placebo group had a larger yearly deterioration than reported in the most recent UK CF registry report (<https://www.cysticfibrosis.org.uk/the-work-we-do/uk-cf-registry/reporting-and-resources>). This may be related to data collection in the registry being from the best time point within a year. We considered whether subjects made efforts to achieve their best FEV₁% in order to fulfil inclusion criteria for trial entry and then regressed to their mean, in support of this, a similar rate of decline on placebo has been shown in one other clinical trial [44]. However, it has not been seen in some others [35, 94]. Regression to the mean could be a plausible answer since the trial incorporated all of the exacerbations and natural progression of CF disease throughout the trial and this may outweigh any superficial improvements for trial entry, particularly as a patient only had to be 2 weeks symptom free. Without assessing pre and post trial data more thoroughly this is difficult to offer as a divisive conclusion and does not explain the slower decline in the active group.

Studies using FEV₁% as an outcome measure need to be of long duration, large numbers and will therefore be expensive to show treatment effects [56, 260]. The European Medicines Agency highlight that a meaningful clinically important difference has not been identified for FEV₁%. So even though the result seen here was modest, any significant difference between treatment and placebo could be clinically relevant. Even an intervention which did not lead to acute improvements but reduced the long-term rate of decline in FEV₁ would be considered by many to be useful, although such studies require longer-term follow up and data most commonly arise from post-trial, phase 4 data collection studies.

(http://www.ema.europa.eu/docs/en_GB/document_library/Report/2012/12/WC500136159.pdf). Sustained benefit of a treatment and changing the decline over time would be a useful treatment effect; but any potential sustained benefit of gene therapy in this study was not reported or indeed powered for at the beginning of the trial.

Further investigation by the GTC is on-going [256] but further proof of concept has only renewed interest in targeting the basic defect for clinical benefit in CF patients.

5.9.4 What can we learn from the MDT about LCI as a clinical trial outcome?

We included LCI in the trial protocol on the basis that it may have provided a more sensitive read-out of improved airway health [148]; in fact, despite an improvement in the primary outcome, FEV₁, LCI response was not different and there was an apparent mismatch

between FEV₁% and LCI results between active and placebo groups. I have therefore attempted to consider possible reasons for this, including whether something about this population or intervention caused the LCI signal to be unresponsive; whether the sample size was suboptimal; whether the study population was not well-suited, having too abnormal baseline LCI and the variability in the assay. An alternative explanation is that LCI is providing the 'true' intervention read-out and the FEV₁ response was somehow spurious; possible explanations for this will be considered.

5.9.4.1 LCI and FEV₁ mismatch

The pre dosing correlation of LCI and FEV₁% is typical of that seen within the literature [87] i.e. a low FEV₁% is indicative of a high LCI (Figure 5.7) and the regression lines of each treatment arm are not different from each other. Interestingly, the post-trial correlations of LCI and FEV₁% diverged (Figure 5.16). Both regression lines were still significant but the treatment groups now differed from each other ($p=0.02$). It is difficult to tell where the differences between groups lie and what the explanations for the divergence could be. As discussed, since the treatment effect on FEV₁% was modest, one explanation for the divergence between groups may be that the placebo groups LCI had deteriorated more than active over the course of the trial. It has also been highlighted that there was a significantly bigger drop in FEV₁ over time in the trial when compared to registry data so the divergence in the graph could reflect this. The fact that only FEV₁%, FVC and gas trapping were statistically significant over trial and the other physiological outcome measures LCI, MEF₂₅₋₇₅ and CT parameters (small mucus plugs and large mucus plugs) did not detect a treatment effect during the trial (caveat that these outcome measures were not powered for) may imply that the treatment was simply not sufficiently effective. Gas trapping had a significant treatment effect during the MDT and small mucus plugs was nearing significance [181]. These CT parameters, as well as large mucus plugs are likely to relate to the conducting airways and small airways disease also highlighted by LCI [74, 264] therefore the use of these measures is justified. All of these outcomes have previously been shown to change with an intervention [122, 265] and as all of the outcome measures continued to correlate with FEV₁% and LCI post trial, it seems likely that the discrepancy between LCI and FEV₁% regression lines post trial seen were in line with natural disease progression. However caution with this exploratory analysis should be applied, particularly where drawing causation from a correlation is likely inappropriate and highlights the dangers of over-interpreting data.

5.9.4.2 LCI is unresponsive to Gene Therapy treatment

This CF population was similar to that in one of the GTC previous studies where an intervention did have a significant impact on LCI [151]; this could be construed as implying that LCI should have been capable of detecting a treatment effect if one were present. However, the CF cohort in our previous study was unstable, experiencing an exacerbation whereas the current trial required stability as an entry criterion; it may be more difficult to detect a change in the latter case.

The limited LCI response seen here in comparison to CFTR targets such as Ivacaftor [38] could be to do with the topical vs. systemic treatment regime. For example with systemic treatment each subject is exposed to the same amount of drug (even adjusted for weight) whereas the nebulisation of a treatment will result in patchy administration due to more poorly ventilated areas of lung. It could also be to do with the more targeted genotype population with ivacaftor treatment and indeed when targeting a larger study population with Orkambi and a larger genotype pool then the treatment response was not as pronounced [40, 94]. LCI as an outcome measure in these small molecule studies was only incorporated for those with mild more preserved lung function, the explanation being that LCI can detect abnormality in mild disease whereas spirometry is limited. Indeed it has been highlighted that lung function is best assessed in early stages by LCI and as lung damage and disease progression continues then FEV₁ becomes more useful [112]. Comparison of LCI in these studies should also be considered with care since the trials used different LCI techniques and as highlighted in Chapter 4, this is not appropriate.

5.9.4.3 Sample size; is LCI underpowered?

Regardless of treatment, the scope of change of FEV₁% and LCI over the time course of the trial may be different. For example the higher the starting FEV₁% the more preserved the lung function so less change in LCI or the lower the FEV₁% the more change in LCI as there is more scope for improvement; but neither seemed to be the case here (Figure 5.18 & 5.19). LCI for both groups changed similarly across the trial so it may be that LCI is a blunted tool in this CF population. For example if the deposition theory is correct and only a specific area of the lung was affected by treatment then LCI may not be precise enough to detect that change due to the nature of test recording whole lung inhomogeneity. If a larger sample size was used the discrepancy between the active group where a flatter regression line was observed compared to the placebo group, may reveal a stabilisation of LCI similar to the FEV₁% results in the main MDT trial but this is speculation.

The lack of significance perhaps relates to the primary outcome only showing a modest change and or the sample size not being powered to show a statistical difference in LCI. However, studies of fewer numbers have been able to show a difference in LCI [152] and in fact post hoc analysis of one study revealed LCI would have been significant with fewer numbers than FEV₁% [38], although this study cohort were patients with FEV₁ above 90% predicted identifying that LCI may only be useful in a milder cohort. The absence of even a non-significant trend in LCI response (even in the milder group) makes this an unlikely explanation in my opinion.

5.9.4.4 Severity

A theory behind the improvement in FEV₁% in the MDT was that those with more severe lung function had increased deposition of pGM169/GL67A in the large airways. If the deposition theory is correct and therefore it is only large airways that have benefitted from increased mucociliary clearance in active treatment, then it may be unsurprising that LCI did not change over the course of the trial. Spirometry reflects large airways and is relatively insensitive to small airway impairment [47, 266] whereas LCI measures heterogeneity across the whole lung but is dominated by poorly ventilated areas in the lung periphery [93]. Phase III analysis is thought to identify differences in gas mixing heterogeneity from the large convective airways vs. the acinar diffusive airways [49] but this analysis has not shown much clinical utility so far [135] and may be limited in more severe disease [121]. Although an updated phase III analysis for more diseased patients is available [114], as the FEV₁% improvement was modest, this time consuming analysis was not undertaken.

The MDT population were older and more severe than some of the other studies utilising LCI (Table 4.1) and baseline and post dosing LCIs in both groups were all above the ULN for LCI, in line with other studies where abnormalities in LCI occur at a young age [1]. The European Medicines agency states that appropriate endpoint selection should be based upon the study population and treatment target. Therefore selection of LCI as an outcome measure in the MDT may have been inappropriate since the cohort was all above the ULN for LCI. However, it seems that those with more severe disease could benefit from using LCI as an outcome measure as there is still scope for change after an intervention (Figure 5.9). Indeed LCI can range from mildly abnormal to grossly increased values for the same FEV₁% so inhomogeneity in advanced disease could be missed by FEV₁% [142]. However in the severe cohort, LCI and FEV₁% are correlated (Figure 5.6). This may be the first indication that LCI may be a blunted tool in more severe CF i.e. the advantage over FEV₁% is not as distinct as

FEV₁% and LCI are correlated and FEV₁% alone can reveal lung function abnormality. The hypothesis that the utility of LCI is limited in severe disease cannot be either confirmed or denied as there may be differences between severity groups but responses to treatment in different severities were exploratory and not powered for.

Interestingly, in those with lung function <70% FEV₁, the r^2 , which describes the spread of data around the slope, are higher post-trial in active but lower post-trial in placebo. This could describe the simultaneous divergence in active and placebo groups by the active treatment making people more variable and the placebo group reverting to the mean value and becoming more stable. This may also be shown in Figure 5.20 where active treatment seemed to promote more movement between quartiles both in a negative and positive direction of change, denoting that those people who had improvements/deterioration in FEV₁% did not change in the same direction for LCI. This may explain the mismatch in FEV₁ data with other outcome measures seen throughout the trial although the data plot was noisy and comparison of the regression lines did not reach statistical significance, so this finding should be interpreted with caution.

It is difficult to draw any specific conclusions from these results, particularly as the overall treatment effect was modest, and the results from smaller subgroups are likely derived by chance. I consider the deposition theory to be the most likely but this would only be clarified with other studies designed to detect differences between severities.

5.9.4.5 Variability of LCI response and natural disease progression

Other interventional studies where LCI has not shown a change after an intervention include short-term chest physiotherapy studies [122, 158, 163] where the explanation for the lack of response was the unpredictable and variable nature of the LCI as well as an influence of previously poorly ventilated airways leading to increased LCI values. The latter could be true in that LCI did increase post trial, therefore the heterogeneous treatment response, even in FEV₁% (Figure 5.3 B & C), could be an explanation for the lack of LCI response in the trial (Figure 5.5).

The absolute LCI treatment effect seen here was -0.28 (-0.71 to 0.14) (p=0.187). Similar results have been shown previously (Table 1.2 in the introductory section) but LCI has also shown to improve by a large magnitude as well. Overall LCI improved in 21 (34%) subjects in the active group and 14 (26%) people in the placebo (Figure 5.15). It seems LCI may be more changeable (positive or negative) in some people compared to others and the response is independent of the gene therapy treatment. Knowledge about the variability of LCI is

essential to establish a minimal clinically important difference in LCI. Steps have been taken to determine acceptable within test variability [101]. As discussed, the large deterioration over the course of the trial may be because patients had achieved artificial wellness to gain entry into the trial. There were no differences in LCI CV% across the trial suggesting this is not the case but whether or not within test LCI CV% is useful in defining 'stability' is unknown. The clinimetric properties of LCI during stable visits have been shown in adults and children by high repeatability and low within test variability [142]. At screening our results fell within acceptable ranges for variability [101] but this was also the case throughout the trial when exacerbations and apparent clinical instability were incorporated. In preschool subjects the LCI intra-class correlation coefficient was higher in stable visits compared to all visits observed over a 12-month observational period [259]. The authors suggest this relates to some of the within test variability of LCI incorporating variation of symptoms or clinical status. The variability in FEV₁% has also been shown to be an important marker of adherence in asthma and tells researchers about those patients at a higher risk of exacerbation [15, 202, 203] so looking further at the variability of LCI may be warranted. There is heterogeneity in LCI responses therefore the utility of LCI in this cohort for active gene therapy treatment may be limited [47, 267].

The lack of LCI response to gene therapy may simply be related to natural disease progression and the treatment effect being smaller than the LCI change over time. Unlike spirometry, literature on LCIs progression over time is scarce. Studies that do show LCI change over time reflect the increase of LCI with age and a significantly increased slope of decline compared to age matched healthy controls. In preschool patients the rate of change seen in LCI is not reflected in FEV₁ Z scores [259] but in older CF patients, FEV₁ also deteriorates with age but not to the same extent as LCI [1]. Both of these studies show the value of LCI in monitoring CF lung disease over time. LCI is abnormal compared to healthy controls, progresses at a faster rate than spirometry and could predict subsequent decline with age. Another repeated measures study linked LCI with clinically meaningful endpoints (CFQR QoL and Pulmonary exacerbations) and increased LCI's validity as a surrogate outcome measure [87]. More longitudinal repeated measurement studies should be completed to further define the clinimetric properties of LCI, relate change to other predictors of deterioration and how LCI can be utilised more readily.

5.9.4.6 LCI results at an earlier time point

Viewing LCI results at an earlier time point during the trial revealed statistically significant differences between active and placebo groups but this was not the case in FEV₁%, although there was trend for higher FEV₁% in the active group. When splitting subjects by baseline severity (above or below 70% FEV₁), an LCI difference persisted in those with mild lung disease but this was not the case for more severe disease or FEV₁% for both severity categories. This may be the first evidence that active treatment was useful in milder patients i.e. the gene therapy was able to disperse into the lower airways and have a beneficial effect but this was only detectable in milder patients at an earlier time point potentially due to natural disease progression being larger than the treatment effect and or increased instability after DO4 masking an LCI response. This instability was explored by looking at LCI CV% and time to first exacerbation. Increased variability and onset of exacerbations later in the trial might mean that the LCI signal is lost due to increased heterogeneity over time. The between test coefficient of variation of LCI was lower before Dose 4 compared to after but this was confirmed to be related to increased visit numbers, and when visit number before and after Dose 4 were matched then the LCI CV% was similar. This suggests that increased variability is to do with increased visit number rather than actual variability post Dose 4. Definition of a pulmonary exacerbation is not easy nor fully standardised [13, 268]; here, a pulmonary exacerbation was pragmatically defined as administration of a new course of oral or intravenous antibiotics. However the decision to start the treatment was solely with the different clinical teams for each patient therefore the reasoning behind the treatment may vary. It may be that using prescribed antibiotics, as an outcome measure alone does not reflect true instability i.e. planned 3 monthly antibiotic cycles would be incorporated into the results, as would perhaps a treatment course ahead of a holiday or exam period. LCI results before, during and after a pulmonary exacerbation have also been variable [155, 255] (Table 1.2 in the introduction section), with apparent responders and non-responders to treatment. This shows that the recovery from an exacerbation is heterogeneous and LCI may not be the most appropriate tool to evaluate treatment response from a pulmonary exacerbation [259].

It is important to note that the trial was not powered to detect an LCI treatment response at any time point during the trial, let alone an earlier time period so the relevance of the significant LCI data is uncertain and this further analysis is purely exploratory. ANCOVA analysis, as utilised in the published trial data, tends to be more conservative once incorporating the effects of covariates but was not completed in this thesis due to the

complexity of the analysis as well as the already explorative nature of sub dividing results. It may be that a simple t test of the LCI data at this time point was not appropriate and reveals a difference that would not be apparent should full ANCOVA testing be completed.

It is not clear that subjects were truly stable at the beginning of the trial i.e. we did not capture data >4 weeks before the trial so would not necessarily notice a downward trend in stability. The “safe to proceed” document and inclusion/exclusion criteria may only be limited to transient changes in the 4 week stability period and not long term changes or drops seen.

Further investigation into the impact of stability on LCI results should be explored particularly as the sensitivity of an MBW test means that all aspects of gas mixing inhomogeneity are incorporated into the result. There is an apparent earlier detection of change in LCI compared to FEV₁% at Dose 4 but within test variability of LCI and onset of pulmonary exacerbation do not explain why LCI would change at an earlier time point and not at the end of the trial, nor are the results reflective of specific improvements at the end of the trial.

The next two sections incorporated the CFQR QoL to explore whether FEV₁% and LCI results reflected that seen by patient reported outcomes as well as check if any FEV₁% or LCI information could be gained from looking at data before and after the trial.

5.9.5 Patient reported outcomes – CFQR QoL

The CFQR QoL is a patient reported outcome measure validated to show the effects of new treatment, disease progression over time and inform clinical practice [84]. The CFQR QoL has been responsive to treatment [81, 83, 95] but no effect was seen in this trial. As an MCID has been established for CFQR QoL, this was used to benchmark changes in FEV₁% and LCI and see if an MCID for these physiological outcome measures could be proposed.

The lack of correlation between the CFQR QoL respiratory domain and physiological outcome measures is similar to what has been reported previously [81]. The ROC curves provide additional evidence that the physiological parameters LCI and FEV₁% do not relate either significantly or consistently to the patient reported outcome measure, the CFQR respiratory domain. An area under the curve of 0.9-1 would indicate both high sensitivity and specificity of changes in the LCI or FEV₁% to recognise a meaningful change in the respiratory domain but as can be seen here the areas under the curve for both LCI and FEV₁% are low showing that each has poor ability to predict a positive change in the CFQR

score. Both ROC curves cross the “null” line indicating that the (very weak) relationships between LCI and FEV₁% with CFQR improvement are unlikely to be monotonic.

It is likely that the respiratory domain and FEV₁% measure different aspects of clinical responses[81], it could also reflect the modest treatment change/improvement seen in this study and the relatively poor efficacy of LCI and FEV₁% to detect these small changes. In this context, therefore, CFQR QoL could not help to define an MCID in either LCI or FEV₁%.

Patient reported outcomes might be useful in the monitoring of patients over time or response to treatment, particularly as baseline physiology results are increasingly mild. Until further comparison studies using the CFQR and LCI are reported then the results are questionable, this is particularly true, as even the significant FEV₁% result was not correlated to CFQR QoL. The utility of this measure in this context is not currently understood.

5.9.6 Did the MDT change the trajectory of FEV₁ or LCI?

The 34 patients who were also enrolled onto the previous run in study were representative of the whole MDT group population (Table 5.9). This implies that the data used in this section is representative of the whole trial population however the 34 people are only 29% of the whole MDT population so results should be interpreted with care, particularly as results were not weighted according to the number of data points available at each time point. This should be completed if this data were to be explored further.

When monitoring FEV₁% over a 6-year period the only FEV₁% linear regression line to be significant was the placebo group in the follow up period. This result is difficult to interpret as the group was small and increased variability could be apparent due to a longer time period and continuity of testing because the same study team collected all of the data for the duration of the trial but not before or thereafter. The follow up data could also show a negative bias for those who are more unwell as they may frequent a clinic more often than a person who is stable [269] but visit number for both active and placebo groups was similar throughout all of the study periods so this does not seem to be the case. Spirometry has been shown to be fairly repeatable measure over time [1] centres follow spirometry guidelines [54, 58] and staff are adequately trained so the results should be reliable.

A marginal or a mixed model may be a better analysis technique to estimate the average rate of change or different trajectories [270]. These data are also being assessed by the GTC to be reported soon. Accounting for covariates such as age, sex, onset of pseudomonas or genotype would only decrease the effects seen here and these variables would presumably have similar effects on both groups, therefore the active groups supposed stabilisation could

be plausible. Knowledge of FEV₁% and LCI results before and after the MDT does not seem to progress understanding of these measures during the MDT.

5.10 Summary

The multi-dose study of gene therapy, with no specific mutation target, was completed in 2012-2014 [181]. With monthly nebulisation of a functioning CFTR protein vs. placebo, the trial showed a modest improvement or stabilisation in lung function (primary outcome: FEV₁%) in the treatment arm compared to placebo [181]. This suggests that novel therapy targeting the basic defect is realistic and has the potential to prevent the deterioration seen in CF disease.

The MDT provides the first proof of concept that clinical benefit from repeated nebulisation of non-viral CFTR gene therapy can be achieved [181]. Though encouraging, only the primary outcome and a few secondary outcomes achieved statistical significance, the response was only seen in some individuals and supporting data from other outcome measures was inconclusive, although the trial was not powered for these. The half of subjects with lower FEV₁% demonstrated the greatest difference in the primary outcome; the most likely explanation was proximal deposition of the therapeutic product. LCI failed to show a treatment effect, which may link to this hypothesis if insufficient gene therapy was deposited in the distal small airways on which this test is predominantly based. In severely affected patients, this may have been due to mucus plugging and airway obstruction; in milder patients, distal deposition may have occurred, but per unit area of airway at a relatively much lower concentration than required for efficacy. An extensive exploratory analysis of LCI in this patient cohort has been undertaken from which some interesting observations have arisen, but which, overall, does not help clarify this discrepancy. Further trials increasing the dose or treatment regime, teaming the treatment with a potentiator or targeting specific responders may be more likely to succeed. Within the consortium there is also a project pipeline to test a pseudo typed lenti-viral vector to increase the efficiency of gene transfer [39]. Inclusion of LCI as an outcome measure in these and other drug trials will provide a robust data set from which we can attempt more fully to understand the clinimetric properties of this test as a trial outcome measure.

6 Chapter 6 - Can we use the response in LCI to treatments to help define a clinically meaningful change?

6.1 Introduction

As seen in the LCI responses Table (1.2) in the Introduction section (Chapter 1), LCI has been shown to change with interventions but the direction and degree of change varies between treatments and is potentially impacted by disease severity. This section explored how LCI changed after CF patients were started on clinically indicated respiratory treatments prescribed by the clinical team at RBH. All treatment changes were included in the analysis to account for differences in responses and the requirement of a global measure and a minimal clinically important difference (MCID) in LCI that is universal across all interventions. Change in LCI was correlated with FEV₁% and (where available) the patient reported outcome, CFQR QoL.

By knowing an expected or therapeutic LCI treatment change then another clinimetric property of LCI can be identified and verification of LCI as a surrogate outcome measure could be a step closer.

6.2 Hypotheses, Aims and Objectives

Hypothesis 1 - LCI will improve after a treatment intervention and it will correlate with improvements in known measures FEV₁% and CFQR QoL allowing an MCID to be proposed.

Hypothesis 2 - LCI is a more sensitive readout of change in response to treatment interventions than FEV₁%.

Hypothesis 3 – LCI will provide a useful monitoring aid for the study of Allergic Bronchopulmonary Aspergillosis (ABPA)

Aim 1 – Compare LCI, FEV₁ and CFQR QoL results at the start of treatment and at a patient's next visit to see if there is a treatment impact, focussing on three types of intervention:

- a. Those with the intention to 'improve respiratory health'.
- b. Those, which may or may not be expected to lead to a change in pulmonary status (a switch from nebulised antibiotics to a powdered inhaler).
- c. Those on long-term ABPA treatment in whom I considered that longitudinal LCI measures may be informative.

Aim 2 – To seek patients that would most usefully benefit from using LCI as a monitoring tool.

Objective 1 - Study those people who started on the same treatment (within treatment intervention).

Objective 2 – Explore results when split by disease severity (mild ($>70\%$ FEV₁) vs. severe ($<70\%$ FEV₁)).

Objective 3 – Examine LCI in those patients who were considered stable vs. unstable (see below).

Objective 4 – In ABPA, assess LCI with other known biomarkers of ABPA (blood and sputum cultures; specifically Immunoglobulin E (IgE), eosinophil count or cultured aspergillus).

Objective 5 – Assess LCI in those subjects who completed the full protocol i.e. including CFQR QoL.

6.3 Methods

6.3.1 Study subjects and treatment changes

Cystic Fibrosis patients at the Royal Brompton Hospital were approached for this study if they were due to start a new respiratory medication; they were recruited either from out-patient clinics or during a visit for a new inhaled medication ‘challenge’ test (in which the physiotherapist assess bronchoconstriction responses to inhaled drug). Treatments ranged from: the routine commencement of Dornase Alpha (DNase; Pulmozyme) at age 6, new oral antibiotics for acute exacerbations, intravenous antibiotics (IV’s) for scheduled or acute exacerbations requiring a hospital stay, a switch from nebulised antibiotics to powdered inhalers (Colobreathe or Tobipod), new treatment with a CFTR potentiator (Ivacaftor (Kalydeco)) and oral corticosteroids (prednisolone).

6.2.2 Study design

Patients were asked if they would be willing to complete an LCI at the beginning of their treatment (either before starting or in the first two days of starting IV’s on the ward – visit 1) as well as at their next clinic visit (or at the end of their inpatient stay – visit 2) when they were still on the treatment intervention. Some patients with suspected Allergic Bronchopulmonary Aspergillosis (ABPA) or who were already on ABPA treatment were followed up for a longer period of time (from 3 months to 2 years, up to 7 visits) as were those who started ivacaftor treatment (3 visits total) to monitor any longitudinal changes in LCI that may occur. The *ad hoc* nature of recruitment and returning visits, adopted to minimise patient inconvenience meant that the LCI before and after treatment was not collected within a standardised time period. The linking of LCI to a clinic visit also meant that

time was often limited and in some cases only LCI on one device (Exhalyzer D) was completed and/ or only two MBW trials were performed; data shown in the Methods development section had confirmed that the latter did not impact the validity of the results. All recruited patients were split into the groups outlined in the aims and objectives of this chapter (section 6.2).

6.2.3 Other outcome measures

Using the hospital data capture system, spirometry results from the day of the LCI test (or the closest result to the day) were collected as well as other information about the stability of the patient (clinician reported information on wellbeing, exacerbations or reasons for changes to treatment) and other biomarkers (blood and sputum cultures; specifically Immunoglobulin E (IgE), eosinophil count or cultured aspergillus for ABPA monitoring). After three months of recruiting the cystic fibrosis questionnaire revised (CFQR QoL) quality of life questionnaire was approved for use and this was completed before LCI at each visit.

6.2.4 Statistics

LCI results at visit 1 and 2 were compared using paired parametric t tests or non-parametric Wilcoxon signed rank tests depending on sample size and D'Agostino and Pearson normality tests. Where 3 or more visits occurred then repeated measures ANOVAs were used. Other outcome measure results were assessed by t tests from visit 1 and 2 as well as compared to the LCI results using correlation coefficients (Pearson for parametric and Spearman for non-parametric data) and linear regression analysis.

6.4 Results

6.4.1 Demographics

72 patients were recruited into the study; Table 6.1 shows the breakdown of patients into the 3 main groups. 26 people (36%) were excluded from analysis for reasons including: no second repeat visit or no repeat visits whilst still on the treatment (n=17), LCI test exclusion criteria (n=5, LCI failing quality due to results being too variable or there was a leak in trace so there were insufficient MBW traces) and not falling into the three main categories (n= 4 started Tai Chi as an intervention or had an induced sputum test that did not culture any pathogens). Table 6.2 shows the patient demographics for each main group.

Table 6.1 Recruitment numbers per treatment intervention and patient breakdown of the three main groups in the study.

Clinically indicated treatment change	Number recruited	Group
IV antibiotics	9	1. Treatment intervention with anticipated improvement – n=32 (44%)
Hypertonic Saline	1	
Colomycin Nebuliser	4	
DNase	7	
Mannitol	1	
Ivacaftor	6	
Other (prednisolone and azithromycin, co-amoxiclav, oral steroids for wheeze)	4	
Colobreathe	4	2. Switch from nebulised antibiotics to a powdered inhaler – n=7 (10%)
Tobipod	3	
ABPA treatment	7	3. ABPA Treatment – n=7 (10%)
Exclusions	26	Exclusions – n=26 (36%)
Total	72	

Table 6.2 Patient demographic data for each treatment intervention group.

Group 1 (n=32- Anticipated improvement)	Mean $\pm$ SD
Gender (%)	14 Female (44%)
Age years (median (min, max))	18.6 (5.4-62.8)
Height cm	152.4 $\pm$ 22.2
Weight kg (median (min, max))	49.2 (19.2-108.6)
FEV ₁ %, Z score ($\pm$ SD) (GLI)	68.5 $\pm$ 16.9%, -2.34 $\pm$ 1.2
Exhalyzer D LCI ($\pm$ SD)	14.1 $\pm$ 5.3
Group 2 (n=7 Powdered inhalers)	Mean $\pm$ SD
Gender (%)	2 Female (29%)
Age years (median (min, max))	18.1 (12.8-46.9)
Height cm	166.9 $\pm$ 4.2
Weight kg (median (min, max))	61.6 (52.8-73.5)
FEV ₁ %, Z score ($\pm$ SD) (GLI)	75.4 $\pm$ 14.6%, -2 $\pm$ 1.2
Exhalyzer D LCI ($\pm$ SD)	13.7 $\pm$ 2.7
Group 3 (n=7 ABPA)	Mean $\pm$ SD
Gender (%)	3 Female (43%)
Age years (median (min, max))	15.2 (8.8-50.5)
Height cm	155 $\pm$ 9.4
Weight kg (median (min, max))	53.7 (36-68.6)
FEV ₁ %, Z score ($\pm$ SD) (GLI)	70.02 $\pm$ 15.3%, -2.49 $\pm$ 1.1
Exhalyzer D LCI ($\pm$ SD)	13.7 $\pm$ 2.2

6.4.2 Anticipated improvement group

6.4.2.1 FEV₁ and LCI results

This section was completed to see if LCI was capable of reading out a treatment response after a clinically indicated treatment change; it was hoped that these data would add to the clinimetric data on LCI and help in the advancement of approval of LCI as a surrogate endpoint in clinical trials.

For the 32 patients with this type of treatment intervention the median (range) number of days between visits 1 and 2 was 29.5 (3 to 178). The fewest number of days between visits was for IV antibiotics; the maximum number of days was for ivacaftor. It was considered that IV antibiotics might promote a fast response to treatment so this patient remained in the analysis cohort. Ivacaftor has been shown to have lasting effects so the longer duration was also considered to be appropriate for inclusion. There was a significant improvement in FEV₁% after treatment (68.5 $\pm$ 17.1 vs. 74.2 $\pm$ 16.8 p=0.001, Figure 6.1) and this was also the same for FEV₁ Z scores (-2.5 $\pm$ 1.3 vs. -2.1 $\pm$ 1.3 p=0.0003). 24 (75%) people had improved (increased) FEV₁ at visit 2 and 8 people had lower FEV₁ at visit 2.

Although there was a trend for an improved LCI (decrease) this did not reach statistical

significance (14.1 ± 5.3 vs. 13.5 ± 4.9 $p=0.1$) (Figure 6.1). The mean decrease in LCI was 0.6 LCI units. LCI improved (decreased) in 17 (53%) people and worsened (increased) in 15.

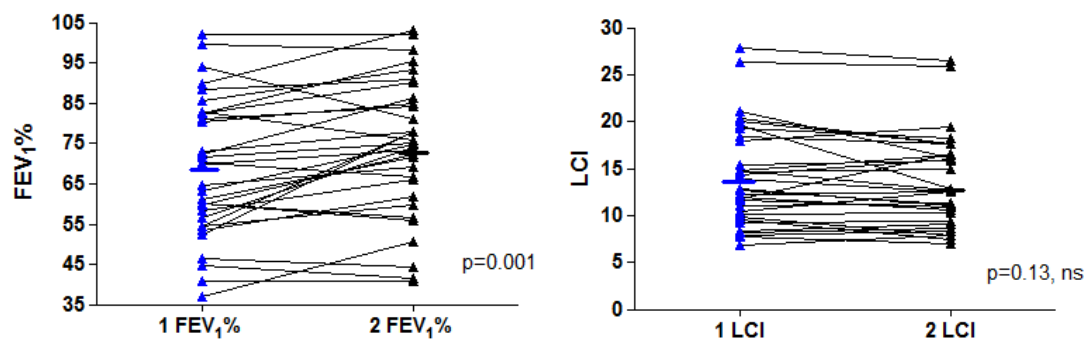


Figure 6.1 FEV₁% and LCI changes from visit 1 (blue) to visit 2 (black) after an intervention. GLI reference ranges for FEV₁%. Note the difference in scales for the y-axes.

Next the delta changes were compared to see if FEV₁ and LCI respond in the same way post treatment (Figure 6.2). For ease, a positive delta was used to reflect improvements in both FEV₁% (FEV₁ increased) and LCI (LCI decreased). There was a significant, but modest, correlation between the two outcome measures ($p=0.0005$).

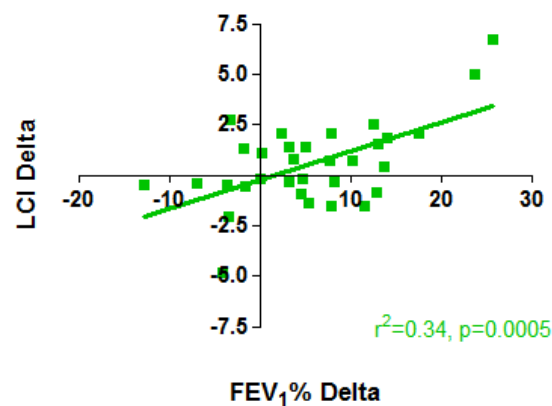


Figure 6.2 Absolute delta changes for FEV₁% and LCI in 32 patients who had a treatment intervention. There was a significant correlation between the changes in the two parameters ($p=0.0005$).

6.4.2.2 Within group changes

The likelihood of some drugs having greater impact than others led to the analysis of visits one and two within treatment group, accepting that numbers for each were small. Table 6.3 shows the results for each group, IV antibiotics and treatment with ivacaftor led to a

significant FEV₁ treatment response, the same was not true for LCI (p=ns). All other groups showed no significant differences in either outcome measure.

Table 6.3 FEV₁ and LCI treatment responses to treatment interventions

Clinically indicated treatment change	Number recruited	FEV ₁ paired t test	LCI paired t test
IV antibiotics	9	p=0.02	p=ns
Hypertonic Saline	1	n/a	n/a
Colomycin Neb	4	p=0.09 (ns)	p=ns
DNase	7	p=ns	p=ns
Mannitol	1	n/a	n/a
Ivacaftor	6	p=0.005	p=ns
Other (prednisolone and azithromycin, co-amoxiclav, oral steroids for wheeze)	4	p=ns	p=ns
Total	32	p=0.0007	p=ns

Ivacaftor FEV₁% and LCI results from both time points as well as a third time point in 4 subjects are shown in Figure 6.3. As mentioned above, from time point 1 to 2, FEV₁% was significantly different (p=0.005); this change also persisted at visit 3 (p=0.04). LCI did not change significantly at either visit.

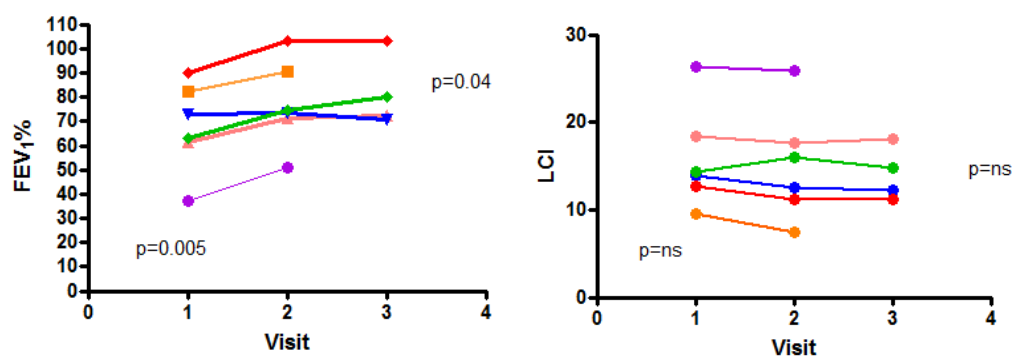


Figure 6.3 FEV₁% and LCI results from those patients (G551D genotype) who started ivacaftor treatment. Different colours represent the same individual.

6.4.3 Severity

The group (32 people) was then split by severity (<70% FEV₁ (17 people) and >70% FEV₁ (15 people) in line with the MDT chapter) to determine whether milder patients were more likely to have a treatment effect that was 'recognised' by LCI (Figure 6.4). FEV₁% showed a significant improvement in those with more severe disease (55.1 ± 8.8 vs. 62.9 ± 12.7 p=0.003) but was not significant in those with mild disease (83.7 ± 9.9 vs. 87.0 ± 10.5 p=0.1).

LCI did not change significantly in either severity group (<70% LCI change was 17.3 ± 5.1 vs. 16.6 ± 4.5 $p=0.3$ and >70%, LCI was 10.5 ± 2.6 vs. 10.1 ± 2.4 $p=0.3$).

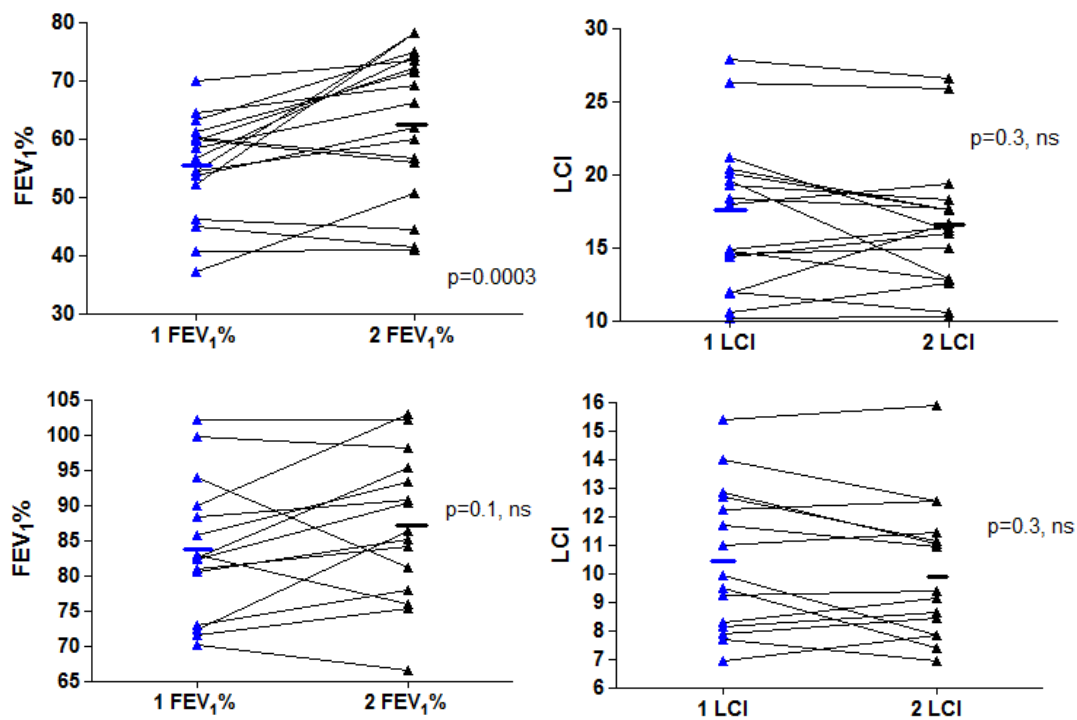


Figure 6.4 FEV₁% and LCI at time point 1 and time point 2 from those with more severe disease at the top and milder patients at the bottom. Note the differences in scales on the y-axes.

6.4.4 Stability

Due to the *ad hoc* nature of recruitment and different treatment interventions, patients will have started treatments for different reasons i.e. experiencing an exacerbation (unstable) or for optimisation respiratory health (whilst stable). Comparison of visits 1 and 2 in stable patients was completed to determine whether the LCI signal is lost in the noise of other factors influencing the stability of a patient.

Stability at visit 1 was based on the reason for starting a new technique and not being unwell at the visit i.e. starting a new clearance technique, starting a new technique for general health optimisation or starting a drug routinely, i.e. DNase at age 6. Stability at visit 2 was improvement after the treatment and documentation in the medical records that the patient was well with no new symptoms. 14 people were considered to be stable at V1 and 14 people were stable at V2, 8 were the same patients.

Of those 8 people stable at both visits, FEV₁% had improved significantly by the second visit but LCI had not (FEV₁% 77.2 ± 13 . vs. 83.4 ± 13.8 $p=0.03$, LCI 10.98 ± 3.1 vs. 10.6 ± 3.3 $p=ns$).

6.4.5 Results assessed alongside CFQR QoL

An MCID is established for the CFQR QoL; in this analysis, I explored whether those people who improved above the MCID in CFQR QoL also improved in FEV₁ and LCI.

Half of the 32 patients with visits before and after a treatment intervention underwent the full protocol including all outcome measures i.e. LCI, FEV₁% and CFQR QoL. Other than the addition of CFQR QoL after ethical approval, the protocol was the same for this group as the treatment group above. Figure 6.5 shows the FEV₁%, LCI and CFQR QoL results for the whole group of these patients: change in FEV₁% and CFQR were significant and LCI showed a trend (FEV₁% 65.6 ± 16.8 vs. 72.1 ± 17.0 p=0.002; CFQR QoL 64.2 ± 17.2 vs. 78 ± 15.2 p=0.003; LCI 15.5 ± 5.96 vs. 14.7 ± 5.9 p=0.06). The different colours represent individuals with a specific intervention. Below 70%, CFQR QoL showed a trend (CFQR QoL 61.8 ± 19.3 vs. 74.99 ± 17.6 p=0.07 (ns)). In milder patients, CFQR QoL was significantly different at the second visit (CFQR QoL 66.7 ± 15.7 vs. 81.9 ± 12.5 p=0.03). There were no significant correlations between change in CFQR QoL, FEV₁% and LCI.

12 out of 16 patients (75%) had a positive change in CFQR QoL above the MCID; of those, 58% of people had positive improvements in LCI and 92% had positive improvements in FEV₁ (Figure 6.6).

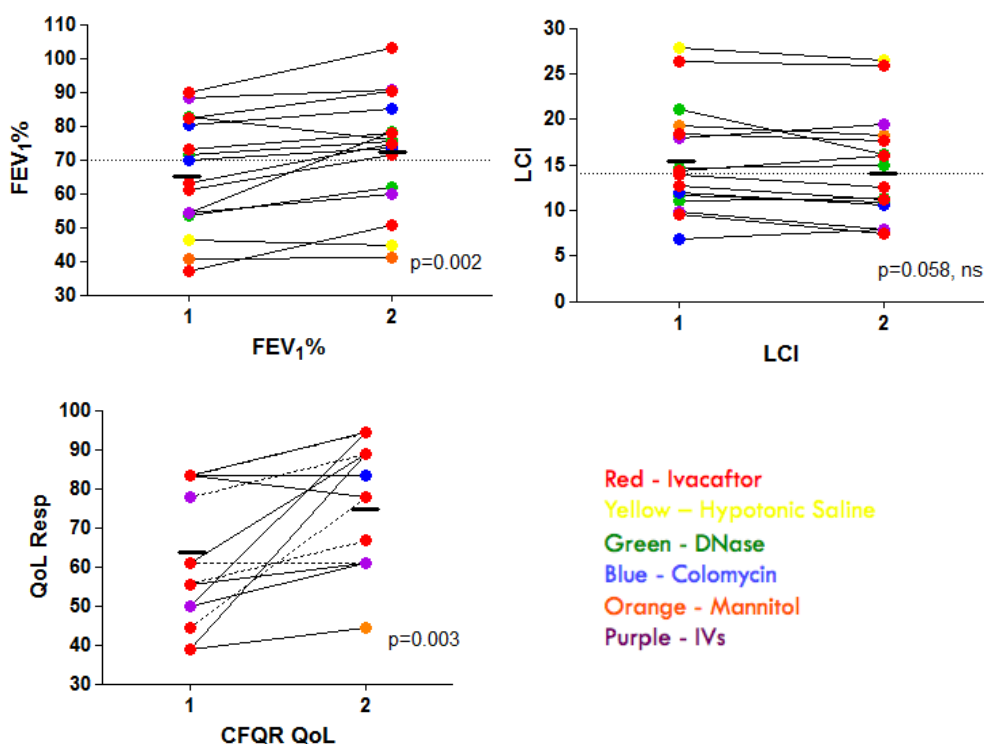


Figure 6.5 FEV₁% (top left), LCI (top right) and CFQR QoL (bottom left) results in 16 patients with a treatment intervention. Both FEV₁% and CFQR QoL were significantly different at the second visit; LCI was nearing significance. The colours represent discrete subjects with specific treatment interventions (see key), dashed lines represent a 70% cut off for severity analysis (mild and severe); the dashed lines between visits on the CFQR graph show those with FEV₁% above 70%. 3 pairs of people had exactly the same CFQR QoL results at visit 1 and 2 that is why some points appear to be missing.

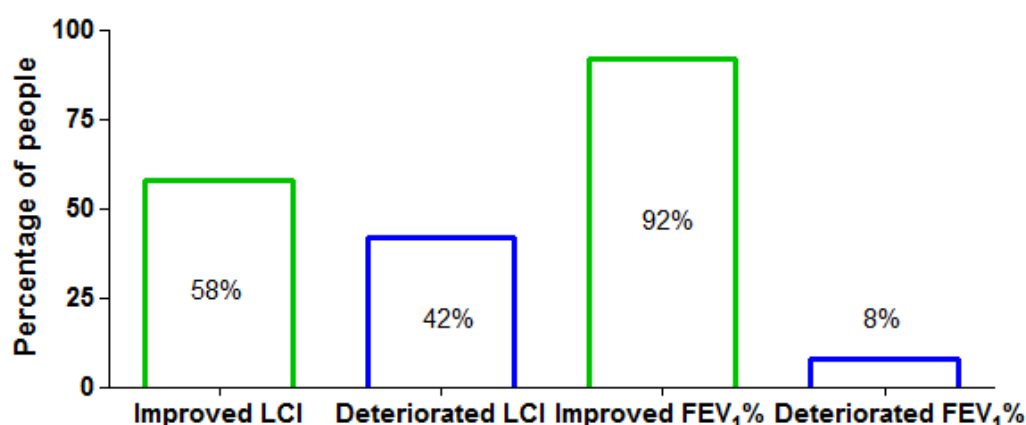


Figure 6.6 Percentage of people with improvements (green) or deterioration (blue) in LCI and FEV₁ in those patients with improved CFQR QoL above the MCID.

6.4.6 Dry powder inhalers

4 people started Colobreathe and 3 people started Tobipod during this study (Table 6.1); they switched onto these inhalers for convenience purposes having been on nebulised formulations of the respective drug, therefore it was not expected that there would be a significant treatment benefit. It was possible that the dry powder drugs could have led to bronchoconstriction and a worsening of FEV₁ and/ or LCI; here I looked at these responses after treatment.

The median number of days between visits 1 and 2 was 79 (57 to 137). There were no differences in FEV₁% or LCI from visit 1 to visit 2 (FEV₁% 75.4 ± 15.8 vs. 73.5 ± 11.5 p=ns; LCI 13.7 ± 2.9 vs. 13.7 ± 2.7 p=ns).

6.4.7 ABPA

ABPA is relatively difficult to diagnose and monitor due to different laboratory and clinical outcome measures used in combination and the fact that patients with ABPA may not show increased symptoms. In this section, I report longitudinal LCI measurements in a small sample of patients with ABPA performed to explore whether LCI offers any additional insight into ABPA monitoring over other measures.

Seven people known to have had previous episodes of ABPA and/or those who were on treatment for ABPA were monitored for up to 2 years. LCI was collected at each visit and all other variables were collected from clinic visits on the same day and or the next closest data capture (FEV₁%, forced vital capacity (FVC), Immunoglobulin (IgE), eosinophil count and cultured *Aspergillus fumigatus*). Three patients with the longest data collection period are presented below (Figure 6.7). Due to the very small numbers, variable visit number and duration period of monitoring, this is an exploratory and preliminary illustration as opposed to a definitive analysis; as such, no formal statistical analysis has been completed.

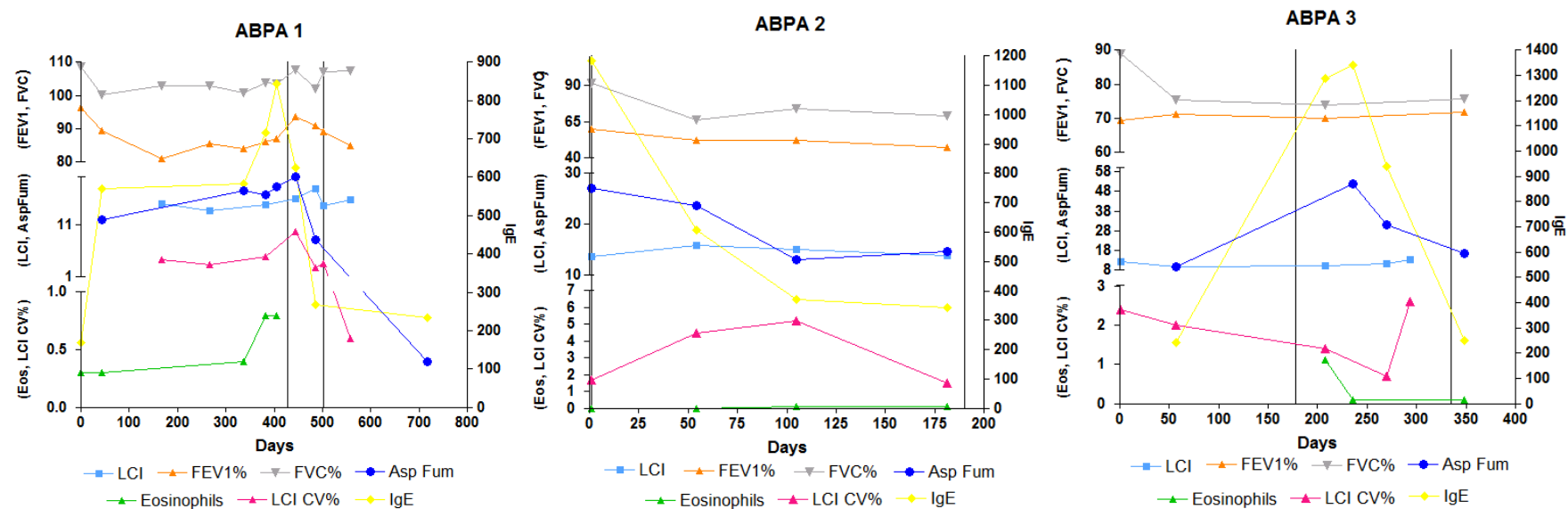


Figure 6.7 Three individual people monitored for ABPA. Days on the x-axis and outcome measures on the y-axes (IgE on the right, FEV₁% and FVC% on the top left axis, LCI and aspergillus on the middle left axis and eosinophil count and LCI CV% on the lower left axis). Note the differences in scales for each individual, completed to show the change of each outcome measure in relation to ABPA treatment (Treatment for ABPA took place between the black solid vertical lines). In all three cases LCI seemed to respond to treatment similarly to FEV₁% and FVC% i.e. with a decrease in FEV₁% or FVC% there was an increase in LCI, LCI did not appear to detect signs of ABPA i.e. show deterioration before FEV₁% and LCI was not necessarily in line or as responsive to treatment than the other clinical biomarkers of ABPA (Immunoglobulin (IgE), eosinophil count and cultured *Aspergillus fumigatus*).

6.5 Discussion

If an MCID for LCI were to be useful then it would need to be universal across all treatment interventions. That, together with the opportunistic approach to recruitment from clinical patients outside an interventional clinical trial, was the rationale behind grouping different treatment interventions together. The ability to capture a large population of participants who had all started on the same treatment within the recruitment period would also have been unrealistic, so grouping interventions facilitated increased patient numbers and the power of the study. FEV₁% showed a significant improvement from time point 1 to time point 2 but LCI did not; the hypothesis that LCI has a more sensitive readout of change in response to treatment interventions than FEV₁ can therefore be rejected.

Unlike the MDT there was a significant but modest delta correlation in FEV₁ and LCI after a treatment, so the outcome measures are responding in the same way but change in LCI was not significant. A fully powered trial designed to assess LCI response to treatment as well as LCI change over time is required. As the absolute LCI change after treatment here was 0.6 LCI units, it is likely that the LCI response needs to be above this to be relevant.

I consider that there are two main explanations for the lack of LCI response; firstly that the timing of the second visit may have occurred before a person had completely recovered/finished a treatment course or after the treatment impact had diminished.

Therefore, as LCI is a sensitive measure of VI, it still detects some of the inhomogeneity from the initial reason for starting the treatment intervention or potentially counts the natural variability or the start of the next exacerbation. Secondly, again because LCI is sensitive, when a treatment occurs, previously non-communicating lung units have been opened that now contribute to the LCI signal. Other factors may be that the interventions are better at targeting large airways rather than peripheral airways or that the latter take longer to respond to treatment. The next sections will explore these points further.

6.5.1 Did the duration of the intervention impact the LCI result?

Table 1.2 in the introduction section reports previous publications of LCI response to treatment; short-term intervention periods with a significant LCI response were from 30 minutes of physiotherapy [164] to 14 days of IV antibiotics [136, 151, 155] and longer term treatment interventions ranged from 4 weeks [38, 152, 159] to a year [135]. In my cohort, the second visit occurred after a median 29.5 days of an intervention but ranged from 3 to 178 days. The optimal time period at which to assess an intervention may depend on a number of factors including the efficacy of the treatment and its mechanism of action.

Therefore, it could be that the treatments here did not have sufficient time to take effect. IVs and ivacaftor treatment interventions made up 47% of the group, which have noted treatment effects after just 14-15 days [38, 151] so the intervention period should have been long enough to see an improvement in LCI. However, LCI did not change significantly even when looking at individual groups, although these will have been underpowered due to small numbers (Table 6.3). As LCI is not a routine clinical test, the lack of response after treatment could suggest that treatment is only continued until FEV₁ has shown an improvement. The researcher did not have an influence on the clinical treatment and or when the patient returned for their second visit however and so this seems unlikely. The treatment duration may also have been too long a duration to see the initial impact of a treatment intervention as natural disease progression may have occurred. FEV₁ was significantly different after treatment so this theory may not apply and the difference could be to do with the sensitivity of LCI.

6.5.2 Is LCI ‘too sensitive’ to demonstrate an improvement after treatment?

The heterogeneous response in LCI can be seen as only half of the individuals had an improvement in LCI compared to 75% who showed FEV₁ improvements. LCI is sensitive to gas mixing inhomogeneity and is dominated by poorly ventilated small airways. This could mean that the treatments had not provided their full effects i.e. whilst changes were occurring, increased inhomogeneity and unevenness in gas mixing were apparent, especially if previously poorly ventilated/non communicating lung units were now contributing to the LCI signal. As FEV₁ improved it suggests that treatment did have an effect so the increased inhomogeneity within the lung may be the most likely explanation for the lack of LCI response.

Reversibility of small airways disease and therefore LCI could also be a factor i.e. are the large airways able to respond to treatment whereas in severe disease the damage to small airways may be irreversible therefore improvements after treatment are not detectable by LCI. Indeed Yammine *et al.* 2014 suggested that the magnitude of reversibility of each component contributing to the washout may well influence the usability of LCI as an outcome measure, particularly with different disease severities [157]. So as LCI is dominated by the small airways, there may not be a signal seen here.

Baseline severity should not influence the ability of an outcome measure to change to a significant degree but all of the patients here had an abnormal LCI so this may play a role in the lack of LCI treatment response. A third test where full treatment response had occurred

and patients were considered stable would have been useful to explore these factors further.

6.5.3 Do drugs with different mechanisms of action differentially affect LCI?

When looking independently at those patients who received ivacaftor (Figure 6.3), there was a significant change in FEV₁ but the same was not true for LCI. This is surprising because of the magnitude of LCI response to ivacaftor seen previously [35, 38]. Amin *et al.* 2011 suggest that FEV₁ and LCI may well respond differently to treatments with different mechanisms of action. For example LCI improved more when hypertonic saline was utilised compared to DNase in a similar 4-week treatment regime. The authors suggest it may be because LCI is more sensitive to treatments that target basic CF pathophysiology i.e. increasing surface hydration rather than further downstream targets of resultant CF disease i.e. a mucolytic agents decreasing the viscosity of mucus in the lungs. While both treatments promote mucociliary clearance, the mechanism of action is different. The ivacaftor results here do not support this hypothesis, although larger more robust studies did show LCI improvements with ivacaftor [38, 94]. These studies were conducted on milder patients where spirometry was within the normal range yet LCI was still able to improve significantly, suggesting widespread improvements to the lung periphery in these patients. LCI may be most useful in this early disease stage and the cohort in this thesis may have been too severe to demonstrate LCI improvements. It also shows that systemic targeted treatment may well promote a larger LCI response compared to other treatments. The *ad hoc*, real life nature of recruitment, that prioritised convenience for the patients, meant that formal power and sample size calculations were not possible for this study, which may well contribute to the lack of response seen here.

Different thresholds of change for different interventions (drugs targeting the downstream effects of CF disease vs. drugs targeting the basic defect) may be a reason for the lack of an MCID for FEV₁ currently. For example 48 weeks of hypertonic saline treatment vs. placebo only led to a 3.2% improvement in FEV₁ [30] whereas DNase resulted in a 5.8% improvement after 24 weeks [29] and Tobramycin a 12% improvement after just 20 weeks [96]. Novel interventions targeting the basic CFTR defect resulted in remarkable changes to FEV₁, 17.2% relative improvement after 24 weeks and significant improvements were even found after just 15 days [35]. With this data in mind it may be unrealistic that identifying an MCID for LCI is possible; it also highlights the number of factors influencing treatment responses. Other recognised outcomes i.e. fewer pulmonary exacerbations or delayed onset of respiratory

signs and symptoms are also still considered to be important responses to treatment, so absolute changes in physiological outcome measures may not be most appropriate final measure. Developing an MCID is a multistep process; if results are so different for physiological outcome measures perhaps the lowest threshold for change should be considered for all treatment targets but the variability of responses also incorporated, particularly as the MCID needs to be larger than the variability seen within the test [52]. For LCI, the clinimetric properties cohesively highlighted by Kent *et al.* 2013 [52] show that LCI is feasible, valid and repeatable so it is a suitable target for the development of an MCID. The results here show that more information is still required about the responsiveness and sensitivity to new and current therapies before a definitive MCID can be determined.

6.5.4 Does severity impact the LCI response to treatment?

FEV₁% significantly improved in the more severe group but this was not the case in the milder disease group (Figure 6.4). This may highlight that FEV₁% does not have a sensitive readout capacity in those with mild disease where spirometry is within the normal range [271, 272] and other outcome measures such as LCI may be more appropriate. However, this was not the case here, where LCI was not significantly different after an intervention in either the mild or severe group (Figure 6.4).

Other studies have also shown that FEV₁% was able to detect an improvement after treatment but LCI was not [122, 156, 157, 160, 163]; these studies and this thesis had participants with a range of disease severities. A physiotherapy study suggests that LCI is less informative than spirometry in those with more severe disease because of significant airflow obstruction and the opening of previously unventilated lung regions [122]. This could be true as short term physiotherapy effects have been described as heterogeneous because of the aim of shifting mucus and opening airways [164]. Similarly, another study suggests that LCI responsiveness is complex and heterogeneity is a result of several components contributing from different areas in the lung. Robinson *et al.* 2009 agree that more distal lung units with slower time constants may start to contribute to the LCI signal after a treatment intervention [155] thus introducing a paradoxical LCI response. These studies suggest that LCI may not be useful in more severe disease.

Yamine *et al.* 2014 completed multivariate regression analysis to determine other physiological factors that could be contributing to the heterogeneous washout responses. They found that FRC minus reserve volume explained 58% of the improved LCI value and, from this, deduced that increased ventilation and reduced hyperinflation were the

mechanisms behind improved LCI [273]. These types of analyses have not been fully explored and utilised, probably related to the complexity of phase III analysis, yet further study of this kind may reveal which aspects, if any, are changing as a result of a treatment.

6.5.5 Does clinical stability influence the LCI response?

It is known that LCI is sensitive, responsive and useful in milder patients [38, 159, 272], but LCI response to treatment is also known to be variable [164]. Stanojevic *et al.* (2016) found that the LCI intra-class correlation coefficient increased when preschool patients were considered to be stable. Intra-class correlation coefficient here describes the proportion of variance between subjects vs. repeated measures, therefore implying that some of the within test variability of LCI depends on patient symptoms and clinical status [259]. The results here show that FEV₁% improved after a treatment intervention but LCI did not. Therefore the stability of each patient was assessed and those considered to be unstable were removed from the dataset to see if the potential increased variability with instability accounted for the lack of LCI response to a treatment where improvement was anticipated. Despite only looking at those people considered to be stable, LCI was still not responsive to treatment, unlike FEV₁% that improved by 6.2%. An abstract looking at stable visits vs. visits where patients were undergoing treatment for a pulmonary exacerbation found that there were statistically different results in both FEV₁ and LCI. Yet, 4 weeks later, after finishing treatment for their pulmonary exacerbation i.e. stable, then the LCI and FEV₁ were not different [274]. Though the numbers are small, this may give an insight into the behaviour of LCI when patients are exacerbating as opposed to the relatively robust and repeatable LCI measure at stable baseline visits. More investigation on stability is required, particularly if treatment responses are transient and or only significant due to the variability of the measure.

Stability sub analysis reduced the cohort to only 8 subjects and 50% of those started DNase for treatment optimisation. DNase has been shown to improve FEV₁ compared to placebo after just two weeks of treatment [29] an effect which persisted for up to two years [275]. The days between DNase treatment visits in the four stable patients here was 14-89 days, FEV₁% was moderate and LCI abnormal at the start. It may be that this is too short a time period to see an LCI benefit but FEV₁ in the large studies of DNase seemed to respond very quickly, even in mild patients.

DNase (pulmozyme) is an enzyme that cleaves extracellular DNA to decrease mucus viscosity and promote airway mucociliary clearance [275]. It could be that the increased mucus

fluidity within the lungs creates too much heterogeneity for an LCI response to occur. For example gas mixing may be more variable because of increased mucus and resultant clearance techniques, and this may be evident from variable LCI responses with physiotherapy [122]. Larger patient samples would likely detect actual DNase responses rather than incorporated variability, indeed a more recent study of LCI response to DNase did show a significant response compared to placebo [152]. This study was a blinded crossover design of 4-week treatment periods but the group did highlight some limitations. These included the DNase group having a lower baseline LCI compared to the placebo; therefore the difference between groups may be factor of that difference and or increased deterioration of LCI promoting the change between treatment arms rather than actual improvement to treatment. The group described the LCI response as promising for the inclusion of LCI as a surrogate outcome measure, although they did suggest that further trials to determine responders vs. non-responders to routine clinical therapies should be completed [152].

The study here did not show an LCI improvement after treatment, even when only stable visits were included, however it is likely that the study was underpowered. Larger longitudinal studies would help to determine the variability and influence of stability on LCI as an outcome measure.

6.5.6 CFQR QoL results after treatment

Similar to the MDT results the CFQR QoL respiratory domain had a large capacity for change after a treatment, yet the change in CFQR QoL respiratory domain in this cohort did not correlate with change in either FEV₁% or LCI. This is similar to the results by Quittner *et al.* (2009) where CFQR QoL results were comparable in magnitude to FEV₁ but were modestly correlated [81]. They suggest that physiological and patient reported outcomes measure different aspects of clinical efficacy.

When developing the MCID for CFQR QoL the group looked at change after tobramycin inhalation therapy on both stable and exacerbating patients [81]. They found that the smallest MCID was in the stable cohort and resulted in a 4-point change in the score on the respiratory domain (5.6 when scaled to 100) [81]. 12 of 16 people in my cohort had improved CFQR QoL above the MCID (Figure 6.6) and interestingly there were more subjects with improved FEV₁ than LCI suggesting that improvements in large airways may be more detectable by the patient than improvements in small airways (detected by LCI). Only larger studies with all outcome measures would confirm this further.

In the 16 people who completed the CFQR QoL, change in LCI was nearing significance. Analysis showed that these 16 were not different from the group who did not complete CFQR in terms of recruitment or time of visits. However, due to the timing of the approval of the CFQR QoL for this study, as well as the European approval of ivacaftor for those patients with G551D, there are a larger proportion of people with a CFQR QoL result that had ivacaftor treatment which is known to improve physiological outcome measures to a large degree [35, 38, 41]. Despite this, when looking at the ivacaftor patients alone, LCI did not improve but this is likely to be the lack of power with the small subject number (Figure 6.3). It could also be chance that those who had completed the CFQR QoL did actually have an improvement to treatment, as it seems unlikely that by completing a CFQR QoL, FEV₁% or LCI would be impacted. Various studies with different interventions have utilised the CFQR QoL [75, 82, 181]. One study utilising ivacaftor therapy even recognised that their cohort had CFQR QoL results that were above the MCID [35]. This suggests that for CFQR QoL, perception of change seems to be similar regardless of the treatment intervention and physiological responses.

Whilst CFQR QoL may be a universal tool for determining patient responses to treatment, as highlighted, physiological responses to treatment are not as uniform or necessarily as responsive to treatment. As described previously in the MDT section, the physiological outcome measures completed here do not seem to reflect the same responses as the patient reported outcome CFQR QoL. The very nature of treating a symptom may mean that a subject feels better- a 'placebo effect' [276] so it may not be surprising that physiological outcome measures do not change to the same extent despite being significantly different. When large numbers were studied significant CFQR QoL results were correlated to FEV₁ and treatment changes [82, 83]. These studies looked at treatment for pulmonary exacerbations and the potentiator ivacaftor. Although the treatment for the pulmonary exacerbation was not consistent, there were worse patient reported outcomes when experiencing a pulmonary exacerbation [82]. There were also other improved domains reflecting the systemic mechanism of action for ivacaftor [83] perhaps increasing the likelihood of change in the patient reported outcome measure. It seems that having a larger cohort with a standardised protocol may increase the likelihood of seeing correlated FEV₁ and CFQR QoL results.

Tepper *et al.* 2014 completed a longitudinal cohort study where 40 CF individuals completed a CFQR QoL as well as a CT at two time points within two years of each other [75]. Whilst there were significant associations between CFQR QoL scores and CT scores at each time

point, they did not find significant relationships of change between the two measures despite CT scores for gas trapping and bronchiectasis getting worse over time. This implies that the CFQR QoL is only a valuable short-term measure and indeed it only assesses patient responses within the last two weeks of completing the questionnaire [81]. This could explain the larger variability of responses and baseline results from the CFQR QoL seen here as the duration of the treatments were two weeks or more and there does not seem to be any correlation with disease severity to the reported CFQR QoL outcomes. The same group did complete a large validation study however and found that younger and milder patients did have higher patient reported scores compared to older and more severe groups [84] but there may be variability on an individual level [79].

Multiple factors including a small sample size, a non-standardised approach to follow up visits as well as a variety of different treatment interventions and large range of participants seem to add to the heterogeneity of the CFQR QoL and LCI results. A large study of Lumacaftor and Ivacaftor in 6-11 year old CF patients homozygous for F508del had a subset of patients complete LCI as well as CFQR QoL [94]. Though the correlations between these results are not reported, both LCI and CFQR results improve after treatment, showing that for mild disease, changes in LCI and CFQR can be similar. Patient reported outcome measures and physiological read-outs should be viewed as complementary markers of treatment response.

6.5.7 LCI and FEV₁ after powdered inhalers

Dry powdered inhalers are more convenient, faster and easier to use compared to nebulised antibiotics [277]. They are well tolerated and have been shown to be non-inferior to nebulised anti-pseudomonal therapy [33, 278, 279]. Both studies used change in FEV₁ as a marker of the powdered inhaled efficacy and there were no differences when compared to nebulised therapy. This is why this group were treated separately to the expected treatment benefit group.

In line with the literature, the results here show there was no difference in FEV₁% after starting a dry powdered inhaler compared to the nebulised formulation of the same drug. For the first time, this has also been shown for LCI; starting Tobipod or Colobreathe does not change gas-mixing homogeneity away from that of a nebulised treatment. Performing LCI in a larger study may reveal differences in the dose deposition or dose delivery of powdered inhalers compared to a nebuliser but as the treatment effects are similar in other larger studies it suggests there is no difference to administration routes. The lack of difference may

also be the difficulties of having naïve patients to therapies and therefore studies need to now compare to current standards of care rather than placebo controlled trials [37]. As LCI is a more sensitive measure of early lung disease than spirometry [73] it may be that should powdered inhalers be superior to nebulised treatment then LCI may be able to detect that improvement. Only a systematic cross over study design would be able to truly address this question. Since there was not direct comparison of LCI before and after starting a nebulised treatment or a placebo group then this cannot be addressed here. Powdered inhalers have not shown superiority over nebulised treatment in other outcome measures i.e. pseudomonas isolates or sputum density [278, 279] therefore it is unlikely that the treatment is superior over conventional nebulisers, even using a more sensitive outcome measure.

The main reason that a powdered inhaler might be superior to a nebuliser is the adherence to treatment, as a powdered inhaler is faster to use and may well be more likely to be taken. The main reason for the individuals starting the treatment at RBH was convenience yet their willingness to complete additional testing as part of this protocol, might suggest that they are engaged with their treatment and had been using their nebuliser before starting the dry powder.

The main adverse event associated with starting a dry powdered inhaler is increased cough [279]. At the Brompton, when a patient is starting a new therapy they undergo a challenge test to assess the tolerability of the new treatment for each individual. Here, all of the patients started powdered inhalers so they were obviously able to tolerate the dry powder. Therefore the difference in LCI in those who are able to tolerate vs. those who cannot has not be addressed.

FEV₁% and LCI do not change after starting a dry powdered inhaler, having been on an equivalent nebulised antibiotic. Anti-pseudomonal treatment is useful in both nebuliser and dry powdered formation and should be initiated per individual preference.

6.5.8 Does LCI help to assess ABPA treatment?

Aspergillus fumigatus is a fungus that grows at body temperature; the spores are of a size that they are deposited in the distal airways and it can produce a large number of toxic and allergenic by-products that are detrimental to CF lung disease [280]. Only a small proportion of CF patients develop ABPA yet it can have severe implications due to its intense inflammatory response to aspergillus in the lung [281]. ABPA is characterised by sensitisation to the allergen and a release of specific IgE [282]. However, diagnosis is a

complex comparison of laboratory and clinical outcome measures [281] with or without increased clinical symptoms (RBH Clinical guidelines - <http://www.rbht.nhs.uk/healthprofessionals/clinical-departments/cystic-fibrosis/clinical-cf-guidelines-care-of-children/care-of-children-with-cystic-fibrosis-respiratory-care/aspergillus-and-other-fungi/>) so increased clinical suspicion is likely for early diagnosis [280]. A recent study of ABPA utilising CF registry data showed that although having ABPA meant that subject's FEV₁% was worse than those without ABPA; ABPA had the smallest impact associated with subsequent FEV₁% decline compared to other risk factors (BMI, high risk genotype, being female, having CF related diabetes and having pseudomonas infection) [283]. The authors suggested that FEV₁ may not be sensitive enough to measure the adverse effects of ABPA therefore I considered that LCI may be more appropriate [283].

With only a small sample of patients it is difficult to confirm whether LCI is useful in the monitoring of ABPA. However, in this cohort LCI did not seem to uniformly improve after ABPA treatment; in fact it was more likely to be increased after treatment (Figure 6.7). There was not a structured visit plan for LCI in the monitoring of ABPA in this cohort i.e. a different time period between each visit and different lengths of monitoring. This will likely add to the variability of LCI response. Sensitivity to *Aspergillus fumigatus* and diagnosis of ABPA is however non-uniform, particularly where most of the major and minor diagnostic criteria for ABPA can be positive in the absence of ABPA [280] (RBH clinical guidelines). Therefore an outcome measure that can read through the variability of ABPA is required. As the clinical impact of ABPA on disease progression is not clearly understood [280] a combination of outcome measures and increased clinical suspicion is likely required to diagnose and track ABPA more thoroughly.

Kraemer *et al.* 2006 completed a longitudinal observational study in 122 children and adolescents with CF [284]. Serial annual lung function tests were performed, including FEV₁, FRC from body plethysmography and LCI. The group found that MEF₅₀ and gas trapping best described lung function decline over time; a diagnosis of ABPA was associated with a higher LCI and a faster rate of decline compared to those without it. LCI was able to differentiate between *Aspergillus fumigatus* sensitisation and ABPA, which led the authors to suggest its most useful application, might be in the early stages of lung function deterioration and development of ABPA. This may explain why in this thesis LCI did not offer an additional benefit for tracking ABPA since all of the patients monitored already had confirmed ABPA rather than LCI being used as a diagnostic tool. The ABPA case studies here did not show a

beneficial use of LCI in monitoring but Kraemer *et al.* 2006 had a large dataset studied for a longer period of time. Closer investigation of the acute changes in LCI would be required to identify what change in LCI is indicative of ABPA on an individual level, but it is likely to be used as an outcome measure combined with all other markers for ABPA recognition.

6.6 Summary of treatment intervention section

LCI did not show a significant treatment response when various respiratory interventions were grouped together, nor when small groups were examined individually. It is likely that the study design and *ad hoc* nature of the second visit added to the variability of the response, as LCI may be influenced by the increased inhomogeneity when treatment interventions promote change. The study does highlight the difficulties of patient monitoring with physiological outcomes at clinic visits rather than structured research visits, therefore this variability should be considered when reviewing data. Stability and severity analysis did not seem to influence the results here but may certainly play a role in the heterogeneity of LCI response. LCI may also respond differently to treatments with different mechanisms of action.

Interventional treatments need to be assessed after at least 2 weeks and with a larger study cohort to account for heterogeneous individual LCI responses. It is likely that for an LCI response to be significant it needs to change above 0.6 LCI units. The results here show that more information is still required about the responsiveness and sensitivity to new and current therapies before a definitive MCID for LCI can be determined.

7 Chapter 7 – Overall Summary and Conclusions

7.1 Overarching Conclusions

This thesis adds to our understanding of the clinimetric properties of LCI; it emphasises that LCI values derived from SF₆ and N₂ are different and the differences are of a magnitude, which is clinically significant. The mechanism is likely to be a contribution of tissue N₂ as shown by extended washouts as well as inherent differences in the way gases behave during a washout as shown by simultaneous washout of both gases using the RMS. There are differences between reference range Z scores therefore caution should be taken when utilising this as an output. The data show no diurnal effects on LCI results potentially due to the small airways not being impacted by circadian variation.

The GTC MDT showed a modest treatment effect for FEV₁ after 12 months of nebulised active gene therapy or placebo. The trial provides the first proof of concept that clinical benefit from repeated nebulisation of non-viral CFTR gene therapy can be achieved. LCI did not show a treatment response during the MDT adding to the hypothesis that the deposition of the active treatment was in the larger airways, rather than the distal airways where the LCI signal predominates. Neither an in-depth exploratory analysis of LCI at an earlier time point nor comparison of severity subgroups provided further insight into the lack of overall LCI treatment response to gene therapy. Although considered to be a more sensitive outcome measure than FEV₁ in early stage disease, this was not seen here; additional consideration may need to be given to the route of administration of an intervention for LCI to be a useful read-out of efficacy.

Knowledge of LCI responsiveness is essential to determine the minimal clinically important difference of LCI and promote LCI to surrogate endpoint status. I had hoped to contribute to the identification of an MCID by using LCI around a clinically indicated change in treatment. Somewhat disappointingly my results show that LCI treatment response (when used in a clinical context rather than a rigorously controlled clinical trial) is variable; this may be due to the different mechanisms of action of various interventions, the stability of the patient and/ or severity of baseline disease in the cohort. An MCID could not be determined from this cohort, which was limited by small numbers and a pragmatic approach to visit timing around patient convenience. Hypotheses have been generated around the required duration and nature of interventions suitable for LCI monitoring, as well as a lower limit likely to be measurable; these need to be tested in further prospective studies.

The utility of LCI in the work covered by this thesis was not as high as expected. Each cohort of patients already had abnormal LCI values and sub analysis of severity and stability may

mean the results lacked power to offer explanations for the variability of LCI responses to treatment. However, I have obtained some useful insights into the clinimetric properties of LCI i.e. feasibility and validity and generated novel data describing differences between techniques. These results emphasise that LCI results from different techniques should be reported with equipment and test gas utilised so that misinterpretation of results is avoided.

7.2 The Future for Lung Clearance Index as an outcome measure

Below, I will consider areas for further research and approaches that could be taken to answer outstanding questions:

7.2.1 SF₆ and N₂

Although the European Clinical Trials Network and the North American Therapeutics Development Network have standardised around N₂ washout at the current time, there remain a number of different techniques and test gases in use globally for LCI measurements; indeed, in our own centre we have four different pieces of equipment. My very strong recommendation is that as long as multiple techniques are in use, **results are always reported together with test gas and equipment**. Although challenging, and currently recommended to be ignored by the latest ERS/ATS MBW standardisation document, I consider that seeking to establish **a correction factor for tissue N₂** contribution would be a worthwhile endeavour. The amount of N₂ will likely be dependent on a number of factors which are both difficult to quantify and variable i.e. actual N₂ tissue release from different sources (which may be further influenced by lung inflammation) and cardiac output. Recognition of differences between test gases may serve as an interim solution to the lack of standardisation between devices especially since different techniques are likely to be utilised within different centres for research and trial purposes. Together with the consensus statement I propose that manufacturers should work with clinicians to determine appropriate software algorithms to align and **minimise the differences between MBW systems** as much as possible. Only a study using both devices in parallel and possessing adequate power to detect LCI treatment response will identify the impact of using different devices and **any advantage of one device over another**.

7.2.2 Interventional studies required

Clearly, different interventions may require differing duration of administration before any treatment effect is visible: as examples at each end of the spectrum, salbutamol inhalers administered to an individual with bronchoconstriction might require a matter of minutes;

inhaled antibiotics to treat chronic infection, months. The degree of improvement and its durability may also depend on host features. Given the variability of LCI response I have demonstrated after the various interventions in this thesis, I would propose that future studies in this area include ***repeated, longitudinal LCI testing at protocol-defined time points***. It is important to consider that there is not an MCID available for FEV₁ so the fact that identification of an MCID here was not possible may not be surprising. More information about the variability of LCI and more studies detecting treatment responses to different interventions will provide more information about a value that would be considered a real change in LCI (rather than variability) although whether a definitive MCID for LCI will ever be identified that is applicable to all interventions and for all disease severities, remains to be seen.

7.2.3 Identifying the optimal patient population for added benefit of LCI

LCI should be selected on the basis of added value of the result e.g. showing abnormality in mild disease where FEV₁ is within normal limits, rather than simply as an additional test, particularly where the feasibility (time commitment) is an issue. Most of the work on LCI has been undertaken in young children, so little is known about LCI variability, responsiveness and progression over time in more severe disease, those with ABPA or more complicated pathogenesis; so large studies are required to identify when (if at any point) LCI may no longer provide information that can be beneficial over traditional lung function measures. So far LCI has mainly been utilised as a research tool but at my institute and other specialist centres LCI is increasingly being requested as a clinical tool. Clinical decision-making is unlikely to be based on a LCI result alone and much like any physiological outcome measures it is likely to be used synergistically with other clinical information. Researchers should endeavour to pool their longitudinal clinical data so the knowledge of clinimetric properties of LCI continues to grow.

7.2.4 Understanding the regional ventilation inhomogeneity of LCI

Although the LCI signal is dominated by abnormality in the small airways, LCI shows gas-mixing inhomogeneity throughout the whole lung and is unable to directly identify where any specific abnormality is occurring. There is increased interest in the use of imaging techniques to identify this abnormality. Specifically CT and MRI techniques that integrate MBW protocols have been developed to not only reveal where the abnormality is occurring but also the extent of the damage. I predict that these techniques will only continue to be

developed and provide a useful technique where the location of abnormality is so far lacking in LCI testing.

7.2.5 Standardisation

All of the points above including knowledge about specific devices, requirement for further interventional trials, information about variability and longitudinal monitoring of LCI in CF patients will not be possible without rigorous quality control criteria and standardised guidelines for MBW and LCI testing. The ERS/ATS MBW 2013 consensus statement takes steps to identify specific equipment requirements and provides guidelines on the acceptable variability within testing sessions; it also highlights general acceptability criteria for MBW testing. The ECFS CTN LCI core facility together with the LCI cores in Toronto, Canada and Sydney, Australia have aligned on an MBW standard operating procedure for a specific device, the Exhalyzer D. Together they have trained, qualified and supported sites in MBW testing as well as over-read and analysed MBW results from several interventional studies including a recent successful trial, the first of its kind to use LCI as the primary outcome [94]. This collaborative effort has produced precise, reproducible results that are beginning to support the acceptability of LCI as a surrogate endpoint for clinical trials. Though thorough manuals and validation studies have taken place, other commercial devices should seek to undertake similar standardisation packages so that results between centres are as comparable and utility of a device is widespread.

Finally, for the identification of abnormality in relation to normal respiratory health, a large, standardised healthy control MBW reference range is required. At the moment centres are likely to have their own healthy control dataset but as shown in the method development section differences between sites even in a healthy control population are apparent. It is only with large reference ranges that all of the variability is considered and LCI using the same techniques can be comparable between studies and sites. This reference range should include the calculation of Z scores so that the LCI results are transferrable and offer more robust results incorporating covariates such as age.

Many groups, working together in this area and sharing results is likely to lead to the potential of LCI to be realised most effectively, and for a multicentre dataset to be built allowing a full understanding of LCI's uses and limitations.

8 References

1. Kraemer, R., et al., *Ventilation inhomogeneities in relation to standard lung function in patients with cystic fibrosis*. Am J Respir Crit Care Med, 2005. **171**(4): p. 371-8.
2. Registry, U.K.C.F., *2015 Annual Data Report*. 2015.
3. Farrell, P.M., *The prevalence of cystic fibrosis in the European Union*. J Cyst Fibros, 2008. **7**(5): p. 450-3.
4. Elborn, J.S., et al., *Current strategies for the long-term assessment, monitoring, and management of cystic fibrosis patients treated with CFTR modulator therapy*. J Cyst Fibros, 2016.
5. Rowe, S.M., S. Miller, and E.J. Sorscher, *Cystic fibrosis*. N Engl J Med, 2005. **352**(19): p. 1992-2001.
6. Riordan, J.R., et al., *Identification of the cystic fibrosis gene: cloning and characterization of complementary DNA*. Science, 1989. **245**(4922): p. 1066-73.
7. Armstrong, D.K., et al., *Gene therapy in cystic fibrosis*. Arch Dis Child, 2014.
8. Elborn, J.S., *Cystic fibrosis*. Lancet, 2016.
9. Quinton, P.M., *Cystic fibrosis: a disease in electrolyte transport*. The FASEB Journal, 1990. **4**(10): p. 2709-17.
10. Matsui, H., et al., *Evidence for Periciliary Liquid Layer Depletion, Not Abnormal Ion Composition, in the Pathogenesis of Cystic Fibrosis Airways Disease*. Cell. **95**(7): p. 1005-1015.
11. Williams, O.W., et al., *Airway Mucus: From Production to Secretion*. American Journal of Respiratory Cell and Molecular Biology, 2006. **34**(5): p. 527-536.
12. Quinton, P.M., *Cystic fibrosis: impaired bicarbonate secretion and mucoviscidosis*. Lancet, 2008. **372**(9636): p. 415-7.
13. Rosenfeld, M., et al., *Defining a pulmonary exacerbation in cystic fibrosis*. J Pediatr, 2001. **139**(3): p. 359-65.
14. Ramsey, K.A., et al., *Early respiratory infection is associated with reduced spirometry in children with cystic fibrosis*. Am J Respir Crit Care Med, 2014. **190**(10): p. 1111-6.
15. Morgan, W.J., et al., *Forced Expiratory Volume in 1 Second Variability Helps Identify Patients with Cystic Fibrosis at Risk of Greater Loss of Lung Function*. J Pediatr, 2015.
16. Sanders, D.B., et al., *Failure to Recover to Baseline Pulmonary Function after Cystic Fibrosis Pulmonary Exacerbation*. American Journal of Respiratory and Critical Care Medicine, 2010. **182**(5): p. 627-632.
17. Sanders, D.B., et al., *Return of FEV1 after pulmonary exacerbation in children with cystic fibrosis*. Pediatric Pulmonology, 2010. **45**(2): p. 127-134.
18. Konstan, M.W., et al., *Risk factors for rate of decline in forced expiratory volume in one second in children and adolescents with cystic fibrosis*. J Pediatr, 2007. **151**(2): p. 134-9, 139.e1.
19. Harun, S.N., et al., *A systematic review of studies examining the rate of lung function decline in patients with cystic fibrosis*. Paediatr Respir Rev, 2016.
20. Conole, D. and G.M. Keating, *Colistimethate Sodium Dry Powder for Inhalation: A Review of Its Use in the Treatment of Chronic Pseudomonas aeruginosa Infection in Patients with Cystic Fibrosis*. Drugs, 2014.
21. Pittman, J.E., et al., *Cystic fibrosis: NHLBI Workshop on the Primary Prevention of Chronic Lung Diseases*. Ann Am Thorac Soc, 2014. **11 Suppl 3**: p. S161-8.

22. Farrell, P.M., et al., *Guidelines for diagnosis of cystic fibrosis in newborns through older adults: Cystic Fibrosis Foundation consensus report*. J Pediatr, 2008. **153**(2): p. S4-S14.
23. Farrell, P.M., et al., *Diagnosis of Cystic Fibrosis: Consensus Guidelines from the Cystic Fibrosis Foundation*. J Pediatr, 2017. **181s**: p. S4-S15.e1.
24. Registry., C.F.F.P., *Annual Data Report*, <https://www.cff.org/2014-Annual-Data-Report/>. 2014.
25. Loeve, M., et al., *Chest computed tomography: a validated surrogate endpoint of cystic fibrosis lung disease?* Eur Respir J, 2013. **42**(3): p. 844-57.
26. Liou, T.G., et al., *Sputum biomarkers and the prediction of clinical outcomes in patients with cystic fibrosis*. PLoS One, 2012. **7**(8): p. e42748.
27. Davies, J.C., *Cystic fibrosis in 2016: considerable progress, but much more to do*. Lancet Respir Med, 2016. **4**(12): p. 943-945.
28. Stephenson, A.L., et al., *Longitudinal trends in nutritional status and the relation between lung function and BMI in cystic fibrosis: a population-based cohort study*. Am J Clin Nutr, 2013. **97**(4): p. 872-7.
29. Fuchs, H.J., et al., *Effect of aerosolized recombinant human DNase on exacerbations of respiratory symptoms and on pulmonary function in patients with cystic fibrosis*. The Pulmozyme Study Group. N Engl J Med, 1994. **331**(10): p. 637-42.
30. Elkins, M.R., et al., *A controlled trial of long-term inhaled hypertonic saline in patients with cystic fibrosis*. N Engl J Med, 2006. **354**(3): p. 229-40.
31. De Boeck, K., et al., *Inhaled dry powder mannitol in children with cystic fibrosis: A randomised efficacy and safety trial*. J Cyst Fibros, 2017.
32. Sanders, D.B., et al., *Standardized Treatment of Pulmonary Exacerbations (STOP) study: Observations at the initiation of intravenous antibiotics for cystic fibrosis pulmonary exacerbations*. J Cyst Fibros, 2017.
33. Galeva, I., et al., *Tobramycin inhalation powder manufactured by improved process in cystic fibrosis: the randomized EDIT trial*. Curr Med Res Opin, 2013. **29**(8): p. 947-56.
34. Turnbull, A.R. and J.C. Davies, *New drug developments in the management of cystic fibrosis lung disease*. Expert Opin Pharmacother, 2016: p. 1-10.
35. Ramsey, B.W., et al., *A CFTR potentiator in patients with cystic fibrosis and the G551D mutation*. N Engl J Med, 2011. **365**(18): p. 1663-72.
36. O'Reilly, R. and H.E. Elphick, *Development, clinical utility, and place of ivacaftor in the treatment of cystic fibrosis*. Drug Design, Development and Therapy, 2013. **7**: p. 929-937.
37. Amin, R. and F. Ratjen, *Emerging drugs for cystic fibrosis*. Expert Opin Emerg Drugs, 2014. **19**(1): p. 143-55.
38. Davies, J., et al., *Assessment of clinical response to ivacaftor with lung clearance index in cystic fibrosis patients with a G551D-CFTR mutation and preserved spirometry: a randomised controlled trial*. Lancet Respir Med, 2013. **1**(8): p. 630-8.
39. Griesenbach, U., K.M. Pytel, and E.W. Alton, *Cystic Fibrosis Gene Therapy in the UK and elsewhere*. Hum Gene Ther, 2015.
40. Wainwright, C.E., et al., *Lumacaftor-Ivacaftor in Patients with Cystic Fibrosis Homozygous for Phe508del CFTR*. N Engl J Med, 2015. **373**(3): p. 220-31.
41. Milla, C.E., et al., *Lumacaftor/Ivacaftor in Patients Aged 6-11 Years With Cystic Fibrosis Homozygous for F508del-CFTR*. Am J Respir Crit Care Med, 2016.
42. Rowe, S.M., et al., *Lumacaftor/Ivacaftor Treatment of Patients with Cystic Fibrosis Heterozygous for F508del-CFTR*. Ann Am Thorac Soc, 2016.
43. McKone, E.F., et al., *Effect of genotype on phenotype and mortality in cystic fibrosis: a retrospective cohort study*. The Lancet, 2003. **361**(9370): p. 1671-1676.

44. Kerem, E., et al., *Ataluren for the treatment of nonsense-mutation cystic fibrosis: a randomised, double-blind, placebo-controlled phase 3 trial*. *Lancet Respir Med*, 2014. **2**(7): p. 539-47.
45. Stocks, J., L.P. Thia, and S. Sonnappa, *Evaluation and use of childhood lung function tests in cystic fibrosis*. *Current Opinion in Pulmonary Medicine*, 2012. **18**(6): p. 602-608 10.1097/MCP.0b013e328358dfbe.
46. Weibel, E.R., *Geometry and Dimensions of Airways of Conductive and Transitory Zones*, in *Morphometry of the Human Lung*. 1963, Springer Berlin Heidelberg: Berlin, Heidelberg. p. 110-135.
47. Robinson, P.D., M.D. Goldman, and P.M. Gustafsson, *Inert gas washout: theoretical background and clinical utility in respiratory disease*. *Respiration*, 2009. **78**(3): p. 339-55.
48. Gibson, G.J., *Clinical Tests of Respiratory Function* Third edition ed. 2008, USA: CRC Press Taylor & Francis Group.
49. Verbanck, S. and M. Paiva, *Gas mixing in the airways and airspaces*. *Compr Physiol*, 2011. **1**(2): p. 809-34.
50. Paiva, M. and L.A. Engel, *Theoretical studies of gas mixing and ventilation distribution in the lung*. *Physiological Reviews*, 1987. **67**(3): p. 750-796.
51. Chmiel, J.F., M. Berger, and M.W. Konstan, *The role of inflammation in the pathophysiology of CF lung disease*. *Clin Rev Allergy Immunol*, 2002. **23**(1): p. 5-27.
52. Kent, L., et al., *Lung clearance index: Evidence for use in clinical trials in cystic fibrosis*. *J Cyst Fibros*, 2013.
53. *European Medicines Agency; Report of the workshop on endpoints for cystic fibrosis clinical trials*. http://www.ema.europa.eu/docs/en_GB/document_library/Report/2012/12/WC500136159.pdf, 2012.
54. Miller, M.R., et al., *Standardisation of spirometry*. *Eur Respir J*, 2005. **26**(2): p. 319-38.
55. Taylor-Robinson, D., et al., *Understanding the natural progression in %FEV1 decline in patients with cystic fibrosis: a longitudinal study*. *Thorax*, 2012. **67**(10): p. 860-6.
56. Konstan, M.W., et al., *Design and powering of cystic fibrosis clinical trials using rate of FEV1 decline as an efficacy endpoint*. *J Cyst Fibros*, 2010. **9**(5): p. 332-8.
57. Quanjer, P.H., 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**(6): p. 1324-43.
58. Miller, A. and P.L. Enright, *PFT interpretive strategies: American Thoracic Society/ European Respiratory Society 2005 guideline gaps*. *Respir Care*, 2012. **57**(1): p. 127-33; discussion 133-135.
59. Gustafsson, P.M., P. Aurora, and A. Lindblad, *Evaluation of ventilation maldistribution as an early indicator of lung disease in children with cystic fibrosis*. *Eur Respir J*, 2003. **22**(6): p. 972-9.
60. Quanjer, P.H., et al., *[Lung volumes and forced ventilatory flows. Work Group on Standardization of Respiratory Function Tests. European Community for Coal and Steel. Official position of the European Respiratory Society]*. *Rev Mal Respir*, 1994. **11 Suppl 3**: p. 5-40.
61. Rosenthal, M., et al., *Lung function in white children aged 4 to 19 years: I-- Spirometry*. *Thorax*, 1993. **48**(8): p. 794-802.
62. Quanjer, P.H., et al., *Age- and height-based prediction bias in spirometry reference equations*. *Eur Respir J*, 2012. **40**(1): p. 190-7.
63. Stanojevic, S., et al., *Reference ranges for spirometry across all ages: a new approach*. *Am J Respir Crit Care Med*, 2008. **177**(3): p. 253-60.

64. Wanger, J., et al., *Standardisation of the measurement of lung volumes*. Eur Respir J, 2005. **26**(3): p. 511-22.
65. West, J.B., *Respiratory Physiology 9ed. The Essentials*. 2012, Baltimore: Wolters Kluwer Lippincott Williams and Wilkins. 200.
66. Hankinson, J., J. Stocks, and R. Peslin, *Reproducibility of lung volume measurements*. European Respiratory Journal, 1998. **11**(3): p. 787-790.
67. Ramsey, K.A. and S. Ranganathan, *Interpretation of lung function in infants and young children with cystic fibrosis*. Respirology, 2014. **19**(6): p. 792-9.
68. Brody, A.S., et al., *Computed tomography in the evaluation of cystic fibrosis lung disease*. Am J Respir Crit Care Med, 2005. **172**(10): p. 1246-52.
69. Ramsey, K.A., et al., *Lung Clearance Index and Structural Lung Disease on Computed Tomography in Early Cystic Fibrosis*. Am J Respir Crit Care Med, 2016. **193**(1): p. 60-7.
70. Tiddens, H.A. and T. Rosenow, *What did we learn from two decades of chest computed tomography in cystic fibrosis?* Pediatr Radiol, 2014.
71. Hall, G.L., et al., *Air trapping on chest CT is associated with worse ventilation distribution in infants with cystic fibrosis diagnosed following newborn screening*. PLoS One, 2011. **6**(8): p. e23932.
72. Shah, R.M., et al., *High-resolution CT in the acute exacerbation of cystic fibrosis: evaluation of acute findings, reversibility of those findings, and clinical correlation*. AJR Am J Roentgenol, 1997. **169**(2): p. 375-80.
73. Owens, C.M., et al., *Lung Clearance Index and HRCT are complementary markers of lung abnormalities in young children with CF*. Thorax, 2011. **66**(6): p. 481-8.
74. Gustafsson, P.M., et al., *Multiple-breath inert gas washout and spirometry versus structural lung disease in cystic fibrosis*. Thorax, 2008. **63**(2): p. 129-34.
75. Tepper, L.A., et al., *Tracking CF disease progression with CT and respiratory symptoms in a cohort of children aged 6-19 years*. Pediatr Pulmonol, 2014.
76. Stahl, M., et al., *Comparison of Lung Clearance Index and Magnetic Resonance Imaging for Assessment of Lung Disease in Children With Cystic Fibrosis*. Am J Respir Crit Care Med, 2016.
77. Semple, T., et al., *Current and future approaches to large airways imaging in adults and children*. Clin Radiol, 2017.
78. Theilmann, R.J., et al., *Characterizing Lung Disease in Cystic Fibrosis with Magnetic Resonance Imaging and Airway Physiology*. PLoS One, 2016. **11**(6): p. e0157177.
79. Goss, C.H. and A.L. Quittner, *Patient-reported outcomes in cystic fibrosis*. Proc Am Thorac Soc, 2007. **4**(4): p. 378-86.
80. Goss, C.H., et al., *Patient-reported respiratory symptoms in cystic fibrosis*. J Cyst Fibros, 2009. **8**(4): p. 245-52.
81. Quittner, A.L., et al., *Determination of the minimal clinically important difference scores for the Cystic Fibrosis Questionnaire-Revised respiratory symptom scale in two populations of patients with cystic fibrosis and chronic Pseudomonas aeruginosa airway infection*. Chest, 2009. **135**(6): p. 1610-8.
82. Brumback, L.C., et al., *Pulmonary exacerbations and parent-reported outcomes in children <6 years with cystic fibrosis*. Pediatr Pulmonol, 2014.
83. Quittner, A., et al., *Effect of ivacaftor treatment in patients with cystic fibrosis and the G551D-CFTR mutation: patient-reported outcomes in the STRIVE randomized, controlled trial*. Health Qual Life Outcomes, 2015. **13**: p. 93.
84. Quittner, A.L., et al., *Development and validation of The Cystic Fibrosis Questionnaire in the United States: a health-related quality-of-life measure for cystic fibrosis*. Chest, 2005. **128**(4): p. 2347-54.
85. Ratjen, F., *How Best to Capture Early Cystic Fibrosis Lung Disease?* Am J Respir Crit Care Med, 2017. **195**(3): p. 283-284.

86. Hoo, A.F., et al., *Lung function is abnormal in 3-month-old infants with cystic fibrosis diagnosed by newborn screening*. Thorax, 2012. **67**(10): p. 874-81.
87. Vermeulen, F., et al., *Lung clearance index predicts pulmonary exacerbations in young patients with cystic fibrosis*. Thorax, 2014. **69**(1): p. 39-45.
88. Stick, S., et al., *Early intervention studies in infants and preschool children with cystic fibrosis: are we ready?* Eur Respir J, 2013. **42**(2): p. 527-38.
89. Stanojevic, S. and F. Ratjen, *Physiologic endpoints for clinical studies for cystic fibrosis*. J Cyst Fibros, 2016.
90. Jaeschke, R., J. Singer, and G.H. Guyatt, *Measurement of health status. Ascertaining the minimal clinically important difference*. Control Clin Trials, 1989. **10**(4): p. 407-15.
91. Pellegrino, R., et al., *Interpretative strategies for lung function tests*. European Respiratory Journal, 2005. **26**(5): p. 948-968.
92. *Lung function testing: selection of reference values and interpretative strategies*. American Thoracic Society. Am Rev Respir Dis, 1991. **144**(5): p. 1202-18.
93. Aurora, P., et al., *Multiple breath inert gas washout as a measure of ventilation distribution in children with cystic fibrosis*. Thorax, 2004. **59**(12): p. 1068-73.
94. Ratjen, F., et al., *Efficacy and safety of lumacaftor and ivacaftor in patients aged 6-11 years with cystic fibrosis homozygous for F508del-CFTR: a randomised, placebo-controlled phase 3 trial*. Lancet Respir Med, 2017.
95. Retsch-Bogart, G.Z., et al., *A phase 2 study of aztreonam lysine for inhalation to treat patients with cystic fibrosis and Pseudomonas aeruginosa infection*. Pediatr Pulmonol, 2008. **43**(1): p. 47-58.
96. Ramsey, B.W., et al., *Intermittent administration of inhaled tobramycin in patients with cystic fibrosis*. Cystic Fibrosis Inhaled Tobramycin Study Group. N Engl J Med, 1999. **340**(1): p. 23-30.
97. Heltshe, S.L., et al., *Short-term and long-term response to pulmonary exacerbation treatment in cystic fibrosis*. Thorax, 2016. **71**(3): p. 223-9.
98. Elborn, J.S., et al., *Efficacy and safety of lumacaftor/ivacaftor combination therapy in patients with cystic fibrosis homozygous for Phe508del CFTR by pulmonary function subgroup: a pooled analysis*. Lancet Respir Med, 2016. **4**(8): p. 617-26.
99. De Boeck, K., et al., *Efficacy and safety of ivacaftor in patients with cystic fibrosis and a non-G551D gating mutation*. J Cyst Fibros, 2014. **13**(6): p. 674-80.
100. Aurora, P., et al., *Multiple-breath washout as a marker of lung disease in preschool children with cystic fibrosis*. Am J Respir Crit Care Med, 2005. **171**(3): p. 249-56.
101. Robinson, P.D., et al., *Consensus statement for inert gas washout measurement using multiple- and single- breath tests*. European Respiratory Journal, 2013. **41**(3): p. 507-522.
102. Fowler, W.S., *Lung function studies; the respiratory dead space*. Am J Physiol, 1948. **154**(3): p. 405-16.
103. Robertson, J.S., W.E. Siri, and H.B. Jones, *Lung ventilation patterns determined by analysis of nitrogen elimination rates; use of mass spectrometer as a continuous gas analyzer*. J Clin Invest, 1950. **29**(5): p. 577-90.
104. Becklake, M.R., *A new index of the intrapulmonary mixture of inspired air*. Thorax, 1952. **7**(1): p. 111-6.
105. Horsley, A.R., et al., *Lung clearance index is a sensitive, repeatable and practical measure of airways disease in adults with cystic fibrosis*. Thorax, 2008. **63**(2): p. 135-40.
106. Darling, R.C., A. Cournand, and D.W. Richards, *STUDIES ON THE INTRAPULMONARY MIXTURE OF GASES. III. AN OPEN CIRCUIT METHOD FOR MEASURING RESIDUAL AIR*. J Clin Invest, 1940. **19**(4): p. 609-18.

107. Horsley, A. and J.M. Wild, *Ventilation heterogeneity and the benefits and challenges of multiple breath washout testing in patients with cystic fibrosis*. Paediatr Respir Rev, 2015. **16 Suppl 1**: p. 15-8.
108. Virtue, R.W. and R.H. Weaver, *The mild anesthetic properties of sulfur hexafluoride*. Anesthesiology, 1952. **13**(6): p. 605-7.
109. Dervos, C.T. and P. Vassiliou, *Sulfur hexafluoride (SF6): global environmental effects and toxic byproduct formation*. J Air Waste Manag Assoc, 2000. **50**(1): p. 137-41.
110. Pittman, J.E., et al., *Variability of a closed, rebreathing setup for multiple breath wash-out testing in children*. Pediatr Pulmonol, 2012. **47**(12): p. 1242-50.
111. Horsley, A.R., et al., *Closed circuit rebreathing to achieve inert gas wash-in for multiple breath wash-out*. ERJ Open Research, 2016. **2**(1).
112. Horsley, A. and S. Siddiqui, *Putting lung function and physiology into perspective: cystic fibrosis in adults*. Respirology, 2015. **20**(1): p. 33-45.
113. Horsley, A., *Lung clearance index in the assessment of airways disease*. Respir Med, 2009. **103**(6): p. 793-9.
114. Verbanck, S., et al., *Acinar and conductive ventilation heterogeneity in severe CF lung disease: back to the model*. Respir Physiol Neurobiol, 2013. **188**(2): p. 124-32.
115. Fowler, W.S., E.R. Cornish, Jr., and S.S. Kety, *Lung function studies. VIII. Analysis of alveolar ventilation by pulmonary N2 clearance curves*. J Clin Invest, 1952. **31**(1): p. 40-50.
116. Subbarao, P., et al., *Multiple-Breath Washout as a Lung Function Test in Cystic Fibrosis. A Cystic Fibrosis Foundation Workshop Report*. Ann Am Thorac Soc, 2015. **12**(6): p. 932-9.
117. Subbarao, P., et al., *Corrections to an ATS Workshop Report on Multiple-Breath Washout Testing for Patients with Cystic Fibrosis*. Ann Am Thorac Soc, 2017. **14**(1): p. 145.
118. Jensen, R., et al., *Multiple breath nitrogen washout: a feasible alternative to mass spectrometry*. PLoS One, 2013. **8**(2): p. e56868.
119. Singer, F., et al., *A realistic validation study of a new nitrogen multiple-breath washout system*. PLoS One, 2012. **7**(4): p. e36083.
120. Lum, S., et al., *Age and height dependence of lung clearance index and functional residual capacity*. Eur Respir J, 2013. **41**(6): p. 1371-7.
121. Horsley, A.R., et al., *Effects of cystic fibrosis lung disease on gas mixing indices derived from alveolar slope analysis*. Respir Physiol Neurobiol, 2008. **162**(3): p. 197-203.
122. Pflieger, A., et al., *Short-term effects of physiotherapy on ventilation inhomogeneity in cystic fibrosis patients with a wide range of lung disease severity*. J Cyst Fibros, 2015.
123. Horsley, A., Saunders, C, Gray R et al., *Lung clearance index improves with treatment of an exacerbation in cystic fibrosis*. European Respiratory Journal, 2007. **30**(Supplement 51): p. S450.
124. Sonneveld, N., et al., *Lung clearance index in cystic fibrosis subjects treated for pulmonary exacerbations*. Eur Respir J, 2015.
125. Rosenow, T., et al., *PRAGMA-CF. A Quantitative Structural Lung Disease Computed Tomography Outcome in Young Children with Cystic Fibrosis*. Am J Respir Crit Care Med, 2015. **191**(10): p. 1158-65.
126. Verbanck, S., et al., *Relationships between the lung clearance index and conductive and acinar ventilation heterogeneity*. J Appl Physiol (1985), 2012. **112**(5): p. 782-90.
127. Kraemer, R. and B. Meister, *Fast real-time moment-ratio analysis of multibreath nitrogen washout in children*. J Appl Physiol (1985), 1985. **59**(4): p. 1137-44.

128. Verbanck, S., et al., *Lung clearance index in adult cystic fibrosis patients: the role of convection-dependent lung units*. Eur Respir J, 2013. **42**(2): p. 380-8.
129. Gustafsson, P.M., et al., *Slow and Fast Lung Compartments in Cystic Fibrosis Measured by Nitrogen Multiple-Breath Washout*. J Appl Physiol (1985), 2014.
130. Schibler, A., et al., *Moment ratio analysis of multiple breath nitrogen washout in infants with lung disease*. Eur Respir J, 2000. **15**(6): p. 1094-101.
131. Yammine, S., et al., *Impact of different breathing protocols on multiple-breath washout outcomes in children*. J Cyst Fibros, 2014. **13**(2): p. 190-7.
132. Gustafsson, P.M., *Peripheral airway involvement in CF and asthma compared by inert gas washout*. Pediatr Pulmonol, 2007. **42**(2): p. 168-76.
133. Saunders, C., et al., *Developments in multibreath washout testing in children with cystic fibrosis*. Curr Med Res Opin, 2016: p. 1-18.
134. Aurora, P., W. Kozłowska, and J. Stocks, *Gas mixing efficiency from birth to adulthood measured by multiple-breath washout*. Respir Physiol Neurobiol, 2005. **148**(1-2): p. 125-39.
135. Harman, K., et al., *P206 Changes In Indices Derived From Multibreath Washout (mbw) Following Treatment With Ivacaftor In Patients With Cystic Fibrosis*. Thorax, 2014. **69**(Suppl 2): p. A167.
136. Vanderhelst, E., et al., *Evidence of an Acinar Response following Treatment for Exacerbation in Adult Patients with Cystic Fibrosis*. Respiration, 2014: p. 492-498.
137. Yammine, S., et al., *Feasibility of nitrogen multiple-breath washout in inexperienced children younger than 7 years*. Pediatr Pulmonol, 2016.
138. Robinson, P.D., et al., *Abbreviated multi-breath washout for calculation of lung clearance index*. Pediatric Pulmonology, 2013. **48**(4): p. 336-343.
139. Downing, B., et al., *Feasibility of lung clearance index in a clinical setting in pre-school children*. Eur Respir J, 2016.
140. Herth, F.J., *Multiple Breath Washout Becoming Feasible for the Assessment of Lung Function in Infants and Young Children with Cystic Fibrosis in a Clinical Setting*. Respiration, 2014.
141. Singer, F., et al., *Practicability of nitrogen multiple-breath washout measurements in a pediatric cystic fibrosis outpatient setting*. Pediatr Pulmonol, 2013. **48**(8): p. 739-46.
142. O'Neill, K., et al., *Lung clearance index in adults and children with cystic fibrosis*. Chest, 2016.
143. Verbanck, S., et al., *Ventilation heterogeneity in the acinar and conductive zones of the normal ageing lung*. Thorax, 2012. **67**(9): p. 789-95.
144. Gronbaek, J., et al., *New timesaving predictor-algorithm for Multiple Breath Washout in Adolescents*. Pediatr Res, 2016.
145. Stahl, M., et al., *Multiple Breath Washout Is Feasible in the Clinical Setting and Detects Abnormal Lung Function in Infants and Young Children with Cystic Fibrosis*. Respiration, 2014.
146. Gonem, S., et al., *Validation of a Photoacoustic Gas Analyser for the Measurement of Functional Residual Capacity Using Multiple-Breath Inert Gas Washout*. Respiration, 2014.
147. Irving, S.J., et al., *Lung clearance index and high-resolution computed tomography scores in primary ciliary dyskinesia*. Am J Respir Crit Care Med, 2013. **188**(5): p. 545-9.
148. Ellemunter, H., et al., *Sensitivity of lung clearance index and chest computed tomography in early cf lung disease*. Respiratory Medicine, 2010. **104**(12): p. 1834-1842.

149. Fuchs, S.I., et al., *Lung clearance index: normal values, repeatability, and reproducibility in healthy children and adolescents*. *Pediatr Pulmonol*, 2009. **44**(12): p. 1180-5.
150. Fuchs, S.I., et al., *Feasibility and variability of measuring the Lung Clearance Index in a multi-center setting*. *Pediatr Pulmonol*, 2012. **47**(7): p. 649-57.
151. Horsley, A.R., et al., *Changes in physiological, functional and structural markers of cystic fibrosis lung disease with treatment of a pulmonary exacerbation*. *Thorax*, 2013. **68**(6): p. 532-9.
152. Amin, R., et al., *The effect of dornase alfa on ventilation inhomogeneity in patients with cystic fibrosis*. *Eur Respir J*, 2011. **37**(4): p. 806-12.
153. Subbarao, P., et al., *Lung clearance index as an outcome measure for clinical trials in young children with cystic fibrosis. A pilot study using inhaled hypertonic saline*. *Am J Respir Crit Care Med*, 2013. **188**(4): p. 456-60.
154. Fuchs, S.I. and M. Gappa, *Lung clearance index: clinical and research applications in children*. *Paediatr Respir Rev*, 2011. **12**(4): p. 264-70.
155. Robinson, P.D., et al., *Using index of ventilation to assess response to treatment for acute pulmonary exacerbation in children with cystic fibrosis*. *Pediatr Pulmonol*, 2009. **44**(8): p. 733-42.
156. Welsh, L., et al., *Lung clearance index during hospital admission in school-age children with cystic fibrosis*. *J Cyst Fibros*, 2014. **13**(6): p. 687-91.
157. Yammine, S., et al., *Reasons for heterogeneous change in LCI in children with cystic fibrosis after antibiotic treatment*. *Thorax*, 2014. **69**(2): p. 183.
158. Grillo, L., et al., *The reproducibility and responsiveness of the lung clearance index in bronchiectasis*. *Eur Respir J*, 2015.
159. Amin, R., et al., *Hypertonic saline improves the LCI in paediatric patients with CF with normal lung function*. *Thorax*, 2010. **65**(5): p. 379-83.
160. Amin, R., et al., *A randomized controlled trial to evaluate the lung clearance index as an outcome measure for early phase studies in patients with cystic fibrosis*. *Respir Med*, 2016.
161. Rosenfeld, M., et al., *Inhaled hypertonic saline in infants and children younger than 6 years with cystic fibrosis: the ISIS randomized controlled trial*. *Jama*, 2012. **307**(21): p. 2269-77.
162. Lenherr, N., et al., *Ivacaftor in a young boy with the rare gating mutation S549R - use of lung clearance index to track progress: a case report*. *BMC Pulm Med*, 2015. **15**(1): p. 123.
163. Fuchs, S.I., et al., *Short-term effect of physiotherapy on variability of the lung clearance index in children with cystic fibrosis*. *Pediatr Pulmonol*, 2010. **45**(3): p. 301-6.
164. Grosse-Onnebrink, J., et al., *Chest physiotherapy can affect the lung clearance index in cystic fibrosis patients*. *Pediatr Pulmonol*, 2017.
165. Poncin, W., et al., *Agreement between multiple-breath nitrogen washout systems in children and adults*. *J Cyst Fibros*, 2016.
166. Yammine, S., et al., *Multiple-breath washout measurements can be significantly shortened in children*. *Thorax*, 2013. **68**(6): p. 586-7.
167. Hannon, D., et al., *Shortened Lung Clearance Index is a repeatable and sensitive test in children and adults with cystic fibrosis*. *BMJ Open Respir Res*, 2014. **1**(1): p. e000031.
168. Saunders, C., et al., *WS21.9 Lung clearance index as a clinical trial outcome measure: Establishing the CTN LCI core facility*. *Journal of Cystic Fibrosis*, 2015. **14**, **Supplement 1**: p. S41.

169. Aurora, P., et al., *Lung clearance index at 4 years predicts subsequent lung function in children with cystic fibrosis*. Am J Respir Crit Care Med, 2011. **183**(6): p. 752-8.
170. Singer, F., et al., *Ventilatory response to nitrogen multiple-breath washout in infants*. Pediatr Pulmonol, 2014. **49**(4): p. 342-7.
171. Maiss, M. and I. Levin, *Global increase of SF6 observed in the atmosphere*. Geophysical Research Letters, 1994. **21**(7): p. 569-572.
172. Marozkina, N.V. and B. Gaston, *Nitrogen chemistry and lung physiology*. Annu Rev Physiol, 2015. **77**: p. 431-52.
173. Lundin, G., *Nitrogen elimination during oxygen breathing*. Acta Physiol Scand Suppl, 1953. **111**: p. 130-43.
174. Nielsen, N., J.G. Nielsen, and A.R. Horsley, *Evaluation of the impact of alveolar nitrogen excretion on indices derived from multiple breath nitrogen washout*. PLoS One, 2013. **8**(9): p. e73335.
175. Newth, C.J., P. Enright, and R.L. Johnson, *Multiple-breath nitrogen washout techniques: including measurements with patients on ventilators*. Eur Respir J, 1997. **10**(9): p. 2174-85.
176. Courmand, A., et al., *STUDIES ON INTRAPULMONARY MIXTURE OF GASES. IV. THE SIGNIFICANCE OF THE PULMONARY EMPTYING RATE AND A SIMPLIFIED OPEN CIRCUIT MEASUREMENT OF RESIDUAL AIR*. Journal of Clinical Investigation, 1941. **20**(6): p. 681-689.
177. Yammine, S., et al., *Using the same cut-off for sulfur hexafluoride and nitrogen multiple-breath washout may not be appropriate*. J Appl Physiol (1985), 2015: p. jap.00333.2015.
178. Zuo, L., et al., *Commentaries on Viewpoint: Using the same cut-off for sulfur hexafluoride and nitrogen multiple-breath washout may not be appropriate*. J Appl Physiol (1985), 2015. **119**(12): p. 1513-4.
179. Gerbase, M.W., et al., *Agreement between spirometers: a challenge in the follow-up of patients and populations?* Respiration, 2013. **85**(6): p. 505-14.
180. Stanojevic, S., A. Wade, and J. Stocks, *Reference values for lung function: past, present and future*. Eur Respir J, 2010. **36**(1): p. 12-9.
181. Alton, E.W., et al., *Repeated nebulisation of non-viral CFTR gene therapy in patients with cystic fibrosis: a randomised, double-blind, placebo-controlled, phase 2b trial*. Lancet Respir Med, 2015.
182. Roberts, H.R., et al., *Airflow obstruction in bronchiectasis: correlation between computed tomography features and pulmonary function tests*. Thorax, 2000. **55**(3): p. 198-204.
183. Brody, A.S., et al., *Reproducibility of a scoring system for computed tomography scanning in cystic fibrosis*. J Thorac Imaging, 2006. **21**(1): p. 14-21.
184. Horsley, A., et al., *Enhanced Photoacoustic Gas Analyser Response Time and Impact on Accuracy at Fast Ventilation Rates during Multiple Breath Washout*. PLoS One, 2014. **9**(6): p. e98487.
185. Ahmad, F., et al., *Multiple breath washouts in children can be shortened without compromising quality*. Eur Respir J, 2015. **46**(6): p. 1814-6.
186. Verbanck, S., et al., *Ventilation distribution during histamine provocation*. J Appl Physiol (1985), 1997. **83**(6): p. 1907-16.
187. Benseler, A., et al., *Effect of equipment dead space on multiple breath washout measures*. Respirology, 2015.
188. Haidopoulou, K., et al., *Alveolar LCI vs. standard LCI in detecting early CF lung disease*. Respir Physiol Neurobiol, 2012. **180**(2-3): p. 247-51.
189. Summermatter, S., et al., *Impact of Software Settings on Multiple-Breath Washout Outcomes*. PLoS One, 2015. **10**(7): p. e0132250.

190. Hardaker, K.M., et al., *Is twice the duration of washout sufficient time between multiple breath nitrogen washout tests?* Eur Respir J, 2017. **49**(2).
191. Jensen, R., et al., *A Systematic Approach to Multiple Breath Nitrogen Washout Test Quality.* PLoS One, 2016. **11**(6): p. e0157523.
192. Saunders, C., et al., *134 Training and qualifying international teams in standardised procedures: steps on the learning curve from the CTN LCI Core Facility.* Journal of Cystic Fibrosis, 2016. **15**: p. S85.
193. Raaijmakers, L., et al., *In vitro and in vivo comparison of two multiple breath nitrogen washout devices.* European Respiratory Journal, 2015. **46**(suppl 59).
194. Arieli, R., *Mass spectrometer for respiratory research.* Respir Physiol Neurobiol, 2010. **170**(2): p. 183-4.
195. Kavouridou, C., Gustafsson, P., Rosberg, M. & Lindblad, A., *Comparing two different algorithms of calculating LCI and FRC from N2MBW in patients with CF.* 2015.
196. Brunner, J.X., et al., *Accurate measurement of N2 volumes during N2 washout requires dynamic adjustment of delay time.* J Appl Physiol (1985), 1985. **59**(3): p. 1008-12.
197. Singer, F., et al., *A new double-tracer gas single-breath washout to assess early cystic fibrosis lung disease.* European Respiratory Journal, 2013. **41**(2): p. 339-345.
198. Houtz, B., et al., *Tidal N2 washout ventilation inhomogeneity indices in a reference population aged 7-70 years.* European Respiratory Journal, 2012. **40**(Suppl 56).
199. Medarov, B.I., V.A. Pavlov, and L. Rossoff, *Diurnal variations in human pulmonary function.* Int J Clin Exp Med, 2008. **1**(3): p. 267-73.
200. Hruby, J. and J. Butler, *Variability of routine pulmonary function tests.* Thorax, 1975. **30**(5): p. 548-53.
201. Hlastala, M.P., B. Wranne, and C.J. Lenfant, *Cyclical variations in FRC and other respiratory variables in resting man.* J Appl Physiol, 1973. **34**(5): p. 670-6.
202. Thamrin, C., et al., *Fluctuation analysis of lung function as a predictor of long-term response to beta2-agonists.* Eur Respir J, 2009. **33**(3): p. 486-93.
203. Thamrin, C., et al., *Variability of lung function predicts loss of asthma control following withdrawal of inhaled corticosteroid treatment.* Thorax, 2010. **65**(5): p. 403-8.
204. Calhoun, W.J., *Nocturnal asthma.* Chest, 2003. **123**(3 Suppl): p. 399s-405s.
205. Desjardin, J.A., et al., *Influence of sleep on pulmonary capillary volume in normal and asthmatic subjects.* Am J Respir Crit Care Med, 1995. **152**(1): p. 193-8.
206. Jones, G. and A. Barker, *Reference Intervals.* The Clinical Biochemist Reviews, 2008. **29**(Suppl 1): p. S93-S97.
207. Lum, S., et al., *Lung function in children in relation to ethnicity, physique and socioeconomic factors.* Eur Respir J, 2015.
208. Duong, M., et al., *Global differences in lung function by region (PURE): an international, community-based prospective study.* Lancet Respir Med, 2013. **1**(8): p. 599-609.
209. McCartney, G., et al., *Why do males in Scotland die younger than those in England? Evidence from three prospective cohort studies.* PLoS One, 2012. **7**(7): p. e38860.
210. Stanojevic, S., et al., *The impact of switching to the new global lung function initiative equations on spirometry results in the UK CF registry.* J Cyst Fibros, 2014. **13**(3): p. 319-27.
211. Armstrong, D., et al., *P209 Standardisation Of Lung Clearance Index In A Multicentre Clinical Trial.* Thorax, 2014. **69**(Suppl 2): p. A168-A169.
212. Quanjer, P.H. and S. Stanojevic, *Do the Global Lung Function Initiative 2012 equations fit my population?* European Respiratory Journal, 2016. **48**(6): p. 1782-1785.

213. Jost, K., et al., *Changes in breathing pattern upon 100% oxygen in children at early school age*. Respir Physiol Neurobiol, 2016.
214. Anagnostopoulou, P., Jensen, R., Kranz, N., Yammine, S., Latzin, P., Ratjen. , *New reference values for N2 multiple breath washout outcomes in pre-school and school-aged children*. European Respiratory Journal, 2016. **48**(Supplement 60): p. 371.
215. Engel, L.A., *Gas mixing within the acinus of the lung*. J Appl Physiol Respir Environ Exerc Physiol, 1983. **54**(3): p. 609-18.
216. Stanojevic, S., et al., *Alternative outcomes for the multiple breath washout in children with CF*. Journal of Cystic Fibrosis, 2015. **14**(4): p. 490-496.
217. Welsh, L., C.F. Robertson, and S.C. Ranganathan, *Increased rate of lung function decline in Australian adolescents with cystic fibrosis*. Pediatric Pulmonology, 2013: p. n/a-n/a.
218. Bancalari, E. and J. Clausen, *Pathophysiology of changes in absolute lung volumes*. European Respiratory Journal, 1998. **12**(1): p. 248-258.
219. Tonga, K., et al., *In-vivo and in-vitro functional residual capacity comparisons between multiple breath nitrogen washout devices*. European Respiratory Journal, 2016. **48**(suppl 60).
220. Lundin, G., *Nitrogen elimination from the tissues during oxygen breathing and its relationship to the fat: muscle ratio and the localization of bends*. J Physiol, 1960. **152**: p. 167-75.
221. Sullivan, L., et al., *Nitrogen back-diffusion during multiple-breath washout with 100% oxygen*. Eur Respir J, 2017. **50**(3).
222. Pabary, R., et al., *Does mass spectrometric breath analysis detect Pseudomonas aeruginosa in cystic fibrosis?* Eur Respir J, 2016. **47**(3): p. 994-7.
223. Kharitonov, S., K. Alving, and P. Barnes, *Exhaled and nasal nitric oxide measurements: recommendations*. European Respiratory Journal, 1997. **10**(7): p. 1683-1693.
224. Yammine, S., et al., *Last Word on Viewpoint: Using the same cut-off for sulfur hexafluoride and nitrogen multiple-breath washout may not be appropriate*. Journal of Applied Physiology, 2015. **119**(12): p. 1515-1515.
225. Kane, M., et al., *Correcting for tissue nitrogen excretion in multiple breath washout measurements*. PLoS One, 2017. **12**(10): p. e0185553.
226. Olegård, C., et al., *Estimation of Functional Residual Capacity at the Bedside Using Standard Monitoring Equipment: A Modified Nitrogen Washout/Washin Technique Requiring a Small Change of the Inspired Oxygen Fraction*. Anesthesia & Analgesia, 2005. **101**(1): p. 206-212 10.1213/01.ANE.0000165823.90368.55.
227. Gustafsson, P.M., et al., *The Effect of Inert Gas Choice on Multiple Breath Washout in Healthy Infants - Differences in Lung Function Outcomes and Breathing Pattern*. J Appl Physiol (1985), 2017: p. jap.00524.2017.
228. Latzin, P., et al., *Optimized temperature and deadspace correction improve analysis of multiple breath washout measurements by ultrasonic flowmeter in infants*. Pediatr Pulmonol, 2007. **42**(10): p. 888-97.
229. Shawcross, A., et al., *Accurate lung volume measurements in vitro using a novel inert gas washout method suitable for infants*. Pediatr Pulmonol, 2015.
230. Frey, U., et al., *Specifications for equipment used for infant pulmonary function testing*. ERS/ATS Task Force on Standards for Infant Respiratory Function Testing. European Respiratory Society/ American Thoracic Society. Eur Respir J, 2000. **16**(4): p. 731-40.
231. Vermeulen, F., et al., *Comparison of lung clearance index measured during helium washin and washout in children with cystic fibrosis*. Pediatr Pulmonol, 2013. **48**(10): p. 962-9.

232. Donald P. Schilder, A.R.D.L.F., *Effect of gas density and viscosity on the maximal expiratory flow-volume relationship*. Journal of Clinical Investigation, 1963. **42**(11): p. 1705-1713.
233. Wood, L.D., et al., *Effect of increased gas density on pulmonary gas exchange in man*. J Appl Physiol, 1976. **41**(2): p. 206-10.
234. Johnson, L.R. and H.D. Van Liew, *Use of arterial PO₂ to study convective and diffusive gas mixing in the lungs*. Journal of applied physiology, 1974. **36**(1): p. 91-97.
235. Kentgens, A.C., et al., *Unexpected results in infant multiple breath washout: Is the truth hidden in the washin?* Pediatr Pulmonol, 2017.
236. von Nieding, G., et al., *Simultaneous washout of helium and sulfur hexafluoride in healthy subjects and patients with chronic bronchitis, bronchial asthma, and emphysema*. Am Rev Respir Dis, 1977. **116**(4): p. 649-60.
237. Kicinski, M., *How does under-reporting of negative and inconclusive results affect the false-positive rate in meta-analysis? A simulation study*. BMJ Open, 2014. **4**(8): p. e004831.
238. Gappa, S.F.M., *Helium cannot simply be exchanged for SF₆ in multiple breath washout* European Respiratory Journal, 2009. **34**(Supplement 53): p. 204s.
239. Wang, J.Y., et al., *Lung function measurement with multiple-breath-helium washout system*. Med Eng Phys, 2013. **35**(4): p. 457-69.
240. Robinson, T.D., et al., *The role of hypoventilation and ventilation-perfusion redistribution in oxygen-induced hypercapnia during acute exacerbations of chronic obstructive pulmonary disease*. American Journal of Respiratory and Critical Care Medicine, 2000. **161**(5): p. 1524-1529.
241. Hopkins, S.R., et al., *Ventilation Heterogeneity Measured by Multiple Breath Inert Gas Testing Is Not Affected By Inspired Oxygen Concentration in Healthy Humans*. J Appl Physiol (1985), 2017: p. jap.01013.2016.
242. Petersson, J. and R.W. Glenny, *Gas exchange and ventilation-perfusion relationships in the lung*. European Respiratory Journal, 2014. **44**(4): p. 1023-1041.
243. Strom, M., Houtz, B., Robinson, P., Lindblad, A., Green, K., Nielsen, K., Singer, F., & Gustafsson, P., *Lung Clearance Index (LCI) based on four different alveolar N₂ estimates in a large reference population aged 6-86 years*. 2015.
244. British Thoracic, S., *Guidelines for the measurement of respiratory function*. Respiratory Medicine, 1994. **88**(3): p. 165-194.
245. Naehrlich, L., et al., *Nasal potential difference measurements in diagnosis of cystic fibrosis: an international survey*. J Cyst Fibros, 2014. **13**(1): p. 24-8.
246. Zabner, J., et al., *Adenovirus-mediated gene transfer transiently corrects the chloride transport defect in nasal epithelia of patients with cystic fibrosis*. Cell, 1993. **75**(2): p. 207-16.
247. Crystal, R.G., et al., *Administration of an adenovirus containing the human CFTR cDNA to the respiratory tract of individuals with cystic fibrosis*. Nat Genet, 1994. **8**(1): p. 42-51.
248. Zabner, J., et al., *Repeat administration of an adenovirus vector encoding cystic fibrosis transmembrane conductance regulator to the nasal epithelium of patients with cystic fibrosis*. J Clin Invest, 1996. **97**(6): p. 1504-11.
249. Moss, R.B., et al., *Repeated aerosolized AAV-CFTR for treatment of cystic fibrosis: a randomized placebo-controlled phase 2B trial*. Hum Gene Ther, 2007. **18**(8): p. 726-32.
250. Hyde, S.C., et al., *Repeat administration of DNA/liposomes to the nasal epithelium of patients with cystic fibrosis*. Gene Ther, 2000. **7**(13): p. 1156-65.

251. Alton, E.W., et al., *Cationic lipid-mediated CFTR gene transfer to the lungs and nose of patients with cystic fibrosis: a double-blind placebo-controlled trial*. *Lancet*, 1999. **353**(9157): p. 947-54.
252. Griesenbach, U. and E.W. Alton, *Progress in gene and cell therapy for cystic fibrosis lung disease*. *Curr Pharm Des*, 2012. **18**(5): p. 642-62.
253. McLachlan, G., et al., *Pre-clinical evaluation of three non-viral gene transfer agents for cystic fibrosis after aerosol delivery to the ovine lung*. *Gene Ther*, 2011. **18**(10): p. 996-1005.
254. Alton, E.W., et al., *A Phase I/IIa Safety and Efficacy Study of Nebulized Liposome-mediated Gene Therapy for Cystic Fibrosis Supports a Multidose Trial*. *Am J Respir Crit Care Med*, 2015. **192**(11): p. 1389-92.
255. Horsley, A., Macleod, K, Saunders, C et al., *UK Gene Therapy Consortium Tracking Study: Lung clearance index improves with treatment of an infective exacerbation*. *Pediatric Pulmonology*, 2007. **42**(Supplement 30): p. 335-336.
256. Griesenbach, U., J.C. Davies, and E. Alton, *Cystic fibrosis gene therapy: a mutation-independent treatment*. *Curr Opin Pulm Med*, 2016.
257. Flume, P.A., et al., *Cystic fibrosis pulmonary guidelines: treatment of pulmonary exacerbations*. *Am J Respir Crit Care Med*, 2009. **180**(9): p. 802-8.
258. Tiddens, H.A., et al., *Novel outcome measures for clinical trials in cystic fibrosis*. *Pediatr Pulmonol*, 2014.
259. Stanojevic, S., et al., *Progression of Lung Disease in Preschool Patients with Cystic Fibrosis*. *Am J Respir Crit Care Med*, 2016.
260. Davies, J.C., *Cystic fibrosis: bridging the treatment gap in early childhood*. *Lancet Respir Med*, 2013. **1**(6): p. 433-4.
261. Jones, A.P. and C. Wallis, *Dornase alfa for cystic fibrosis*. *Cochrane Database Syst Rev*, 2010(3): p. Cd001127.
262. Wilms, E.B., et al., *Azithromycin maintenance therapy in patients with cystic fibrosis: a dose advice based on a review of pharmacokinetics, efficacy, and side effects*. *Pediatr Pulmonol*, 2012. **47**(7): p. 658-65.
263. Buzzetti, R. and D. Salvatore, *Combining Clinical Trial and Patient Registry Data in Cystic Fibrosis: Who Should Be Compared?* *Am J Respir Crit Care Med*, 2017. **195**(3): p. 404-405.
264. Ramsey, K.A., et al., *Lung Clearance Index and Structural Lung Disease on Computed Tomography in Early Cystic Fibrosis*. *Am J Respir Crit Care Med*, 2015.
265. Loeve, M., et al., *Reversibility of trapped air on chest computed tomography in cystic fibrosis patients*. *Eur J Radiol*, 2015.
266. Macklem, P.T., *The physiology of small airways*. *Am J Respir Crit Care Med*, 1998. **157**(5 Pt 2): p. S181-3.
267. Robinson, P., Van Asperen P, Cooper, P, et al., *Assessing response to intravenous therapy for pulmonary exacerbations in children with cystic fibrosis*. *Pediatric Pulmonology*, 2006. **41**(Supplement 29): p. 348.
268. Stenbit, A.E. and P.A. Flume, *Pulmonary exacerbations in cystic fibrosis*. *Curr Opin Pulm Med*, 2011. **17**(6): p. 442-7.
269. Sawicki, G.S., et al., *Sustained Benefit from ivacaftor demonstrated by combining clinical trial and cystic fibrosis patient registry data*. *Am J Respir Crit Care Med*, 2015. **192**(7): p. 836-42.
270. Szczesniak, R., et al., *Use of FEV1 in cystic fibrosis epidemiologic studies and clinical trials: A statistical perspective for the clinical researcher*. *J Cyst Fibros*, 2017.
271. Ellemunter, H., et al., *Long-term improvement of lung clearance index in patients with mild cystic fibrosis lung disease: Does hypertonic saline play a role?* *J Cyst Fibros*, 2015.

272. Fuchs, S.I., et al., *Tracking Lung Clearance Index and chest CT in mild cystic fibrosis lung disease over a period of three years*. Respir Med, 2014. **108**(6): p. 865-74.
273. Abbas, C., et al., *Treatment response of airway clearance assessed by single-breath washout in children with cystic fibrosis*. J Cyst Fibros, 2013. **12**(6): p. 567-74.
274. O'Neill, K., Elborn, J, Johnston, E, et al., *Changes in Lung clearance index during periods of stability and pulmonary exacerbation*. Pediatric Pulmonology, 2012. **47**(Supplement 35): p. 352-353.
275. Quan, J.M., et al., *A two-year randomized, placebo-controlled trial of dornase alfa in young patients with cystic fibrosis with mild lung function abnormalities*. J Pediatr, 2001. **139**(6): p. 813-20.
276. Brown, W.A., *The placebo effect in clinical practice*. 2013: Oxford University Press.
277. McKeage, K., *Tobramycin inhalation powder: a review of its use in the treatment of chronic Pseudomonas aeruginosa infection in patients with cystic fibrosis*. Drugs, 2013. **73**(16): p. 1815-27.
278. Konstan, M.W., et al., *Safety, efficacy and convenience of tobramycin inhalation powder in cystic fibrosis patients: The EAGER trial*. J Cyst Fibros, 2011. **10**(1): p. 54-61.
279. Schuster, A., et al., *Safety, efficacy and convenience of colistimethate sodium dry powder for inhalation (Colobreathe DPI) in patients with cystic fibrosis: a randomised study*. Thorax, 2013. **68**(4): p. 344-50.
280. Ohn, M., et al., *Question 11: How should Allergic Bronchopulmonary Aspergillosis [ABPA] be managed in Cystic Fibrosis?* Paediatr Respir Rev, 2016.
281. Hemmann, S., et al., *Differential IgE recognition of recombinant Aspergillus fumigatus allergens by cystic fibrosis patients with allergic bronchopulmonary aspergillosis or Aspergillus allergy*. Eur J Immunol, 1998. **28**(4): p. 1155-60.
282. Thia, L.P. and I.M. Balfour Lynn, *Diagnosing allergic bronchopulmonary aspergillosis in children with cystic fibrosis*. Paediatr Respir Rev, 2009. **10**(1): p. 37-42.
283. Kaditis, A.G., et al., *Effect of allergic bronchopulmonary aspergillosis on FEV1 in children and adolescents with cystic fibrosis: a European Cystic Fibrosis Society Patient Registry analysis*. Arch Dis Child, 2017.
284. Kraemer, R., et al., *Effect of allergic bronchopulmonary aspergillosis on lung function in children with cystic fibrosis*. Am J Respir Crit Care Med, 2006. **174**(11): p. 1211-20.

9 Appendices

9.1 Appendix 1

Methodological steps/issues with simultaneous washout testing

Simultaneous washout was completed to eliminate individual differences as an explanation for the differences between MBW completed on SF₆ or N₂. The final simultaneous washout method is presented in Chapter 4. Here, the methodological steps to the final protocol are described.

Calibrating the Innocor device for use of 100% O₂.

- Initially the Inn/SF₆ device displayed 100% O₂ at only 80% O₂.
- A factory 2-point calibration (Innocor Service Manual, January 2007) was required to increase the detection range through the internal amplifier.
- The Inn/SF₆ could then detect 100% O₂.

Is SF₆ detected on the ExD/N₂?

- The Exhalyzer D system has capabilities to measure SF₆ [197] but the one at RBH does not have this function. The software indirectly calculates the SF₆ signal in much the same way as N₂ i.e. using the O₂ and CO₂ signal and calculating the SF₆ from knowledge of its molar mass signal and the gas fraction after subtracting the O₂ and CO₂ [119]. The studies that use the Exhalyzer D with SF₆ use it at 4-5%.
- When running the 0.2% SF₆ gas mix through the ExD/N₂ it seems the molar mass signal from 0.2% of SF₆ is too low to be detected by the device, as there was no impact on any gas measurement at online analysis. No effects of using SF₆ were found on the ExD/N₂ machine.
- The researcher contacted the manufacturers to see if the software for SF₆ and N₂ could be merged but they are not compatible. Dual software is in development (<http://www.ecomedics.com/index.php?id=29>).
- Although no impact of the 0.2% SF₆ was seen on screen, going forward the researcher was mindful that the 0.2% was likely to be counted as N₂ (the direct measurement of O₂ and CO₂ wouldn't pick up the SF₆ but an indirect calculation of N₂ could incorporate it). This would be considered at analysis.

If sampling from the Exhalyzer D to measure N_2 and breaths, how can SF_6 be measured?

- Drill a hole in the capnostat so that the Inn/ SF_6 sample line could be placed in the flow pass on the ExD/ N_2 ? This was not completed as each capnostat costs £140 (net price – <http://www.annox.co.uk/>) and the result was uncertain.
- There was no obvious sample issue when using a 3-way tap to connect all of the sample lines. However the Inn/ SF_6 is only able to record SF_6 when flow is also recorded on the Innocor. When the participant was breathing on Exhalyzer, the investigator tried to simulate their tidal breaths but unsurprisingly the raw data was not compatible. The raw files (A file from Exhalyzer D and .ino file from Innocor) are displayed with 100 samples per second, therefore they are too large to find and match breaths and end tidal gas concentration (Cet). The 3-way tap may also not be airtight and therefore the likelihood of leaks was high. This method was not adopted.

How to use both devices to measure N_2 and SF_6 at the same time?

- The pieces of equipment that measure flow volume (pnemotach for Inn/ SF_6 and ultrasonic flowmeter for the ExD/ N_2) were attached in series by plastic tubing sourced from the RBH lung function department. The seal was tight and secure.
- The serial order of the devices was tested, but mounting the breathing unit using the original by pass flow block from the ExD/ N_2 (ExD/ N_2 closest to the gas sample and the Inn/ SF_6 closest to the patient) enabled easier testing for subjects (Figure 2.9). See Figure 2.14 for a detailed description of where the devices lay in situ, the deadspace before and after gas concentration recording on both devices and impact to inspiration and expiration values depending on the sample uptake.
- Calibration was performed on both devices with the devices connected to replicate testing conditions.
- Flow calibration was performed with the luer lock in place and CO_2 warmer removed as per normal ExD/ N_2 calibration set up (see methodology section 2.4.4.2). The Inn/ SF_6 was also calibrated as per normal calibration procedure but with the devices connected in series.
- A sample of calibration numbers can be seen in the Table below (Table 1, Appendix 1), for all except channel synchronisation, the values did not seem to vary outside normal calibration values (see methods in Chapter 2).

- Channel synchronisation on the ExD is more sensitive to changes even without adapting the patient interface but the impact during MBW testing was small. Original synchronisation values were returned after this testing.

Table 1 Appendix 1 – Example of calibration data when the devices were connected in series to complete a simultaneous washout.

Exhalyzer D	Sample original values	Sample values with devices attached in series
Environmental settings	Normal	No change
Flow settings (within 2.5%)	Insp deviation = -0.43 Exp deviation = 0.59	Insp deviation = 1.34 Exp deviation = -0.16 No change
Channel settings	Medical air = 20.44 100% O ₂ = 99.8	Medical air = 20.45 100% O ₂ = 99.8 No change
Signal Synchronisation	O ₂ = 0.639 CO ₂ = 0.081 MMss = 0.641	O ₂ = 0.639 CO ₂ = 0.081 MMss = 0.641
Innocor	Sample original values	Sample values with devices attached in series
Environmental settings	Normal	No change
Flow settings (within 0.980 and 1.04)	Fill = 0.978 Empty = 0.995	Fill = 0.985 Empty = 0.999 No change
Gas values (within 40ms from the last test)	O ₂ = 1357 CO ₂ = 1357	O ₂ = 1355 CO ₂ = 1364 No change

Series testing issues

- Three tests were discarded after it was discovered SF₆ would only increase up to 0.16. There was a broken filter connecting the nafion tube to the Innocor causing a leak.
- In isolation the Inn/SF₆ device would record up to 100% O₂, but in some washout's using this set up the O₂ only went up to ~86%. It was likely to do with the order of the gas sampling lines, as the Inn/SF₆ was closest to the mouth, i.e. some of the sample is not recorded due to the deadspace as well as sample uptake from the ExD/N₂ (see figure 2.14). It was unlikely to be a leak as flow was similar between devices and the ExD did read up to 100%. It could have been the inability of the Inn/SF₆ device to measure 100% O₂ accurately, but the manual seems to suggest this should be possible. More investigation was completed but as N₂ was calculated on the ExD/N₂ and this device did not seem to be affected, testing continued. Only

those tests where 100% O₂ was reached on the Inn/SF₆ were used in the final analysis.

- In two cases the raw breath data was not saved on the Exhalyzer D due to a Windows error. The data can be used for the overall comparison but not further.
- Extended washout testing was completed for 10 minutes during the testing phase. The lowest N₂ that could be achieved was 0.9%. The researcher felt slightly lightheaded at this point and as the N₂ result had been constant for at least 4 minutes the washout was not continued further. Two minutes of extended washout was completed in the final testing procedure, a value that was long enough to reach a plateau of low N₂ but also short enough to not induce light headedness.
- So that SF₆ did not accumulate in the testing room, extra tubing was required to feed in and washout the tracer gases (connected at both ends of the bias flow block).
- After the washout of SF₆ and N₂ was completed a further washin of SF₆ and N₂ was recorded. The hypothesis was that if the N₂ washed in quicker or similarly to SF₆ it would show that the differences in washout were to do with potential measurement error in equipment, use of 100% O₂ in washout or contribution from tissue N₂ and not the test gases properties. This is not the first time that analysis of washin has been considered. The use of both washin and washout methods have been suggested to increase the accuracy of results [226]. A recent article has also suggested the degree of washin at the start of a test may have a large impact on the accuracy of the washout result [235]. Washin and washout analysis has also been completed using helium as a tracer gas [231] but washin and washout were considerably different. The authors attributed the differences to methodological sampling of washin and washout, breathing pattern as well as differences between gases mixing in opposite directions. The washout signal is dominated by slowly ventilating lung units (long tail in disease because different lung units have the N₂ diluted at different rates [65]) whereas well-ventilated lung units dominate the washin signal. Difficulties in converting and calculating LCI from a washin signal meant that testing was aborted. If washin and washout results were also different from each other then the hypothesis of this section would not necessarily be answered by this testing.
- Analysis of SF₆ and N₂ during a simultaneous washout was completed using normal software for each device.

Methods: Innocor – 12 people completed a normal room air washout of 0.2% SF₆ vs. a washout of 0.2% SF₆ using 100% O₂. Exhalyzer D – 6 people completed a routine N₂ washout using the Exhalyzer D vs. a washout of N₂ including 0.2% SF₆.

Results: Innocor - There were no differences in tidal volume and respiratory rate between washout gas but LCI and CEV were higher with 100% O₂. FRC was lower with 100% O₂ compared to room air.

Exhalyzer D – There were no differences in any washout parameter when 0.2% SF₆ was being washed out at the same time as N₂.

Conclusions: Software for the Innocor system overestimates LCI when using 100% O₂ when compared to room air washout. Software for the Exhalyzer D does not account for the addition of 0.2% during a washout. Both indicate that the softwares for both devices are not appropriate for the analysis when completing simultaneous MBW testing. Innocor software should be adapted to account for the use of 100% O₂, it may be that the increased viscosity of 100% O₂ leads to confusion in the determination of FRC and CEV, therefore increasing LCI.

The ExD software does not incorporate SF₆ independently from the other gases.

More investigation would be required if this testing were to be used again but ideally specific software, i.e. Testpoint or Matlab would be used, as these have the capabilities to calculate both SF₆ and N₂ washout appropriately. SF₆ software for the ExD is available, yet this uses 4% SF₆. 0.2% SF₆ may be too low a concentration to affect the N₂ washout, yet it is not obvious where the 0.2% SF₆ is being recorded on the ExD software so valid conclusions cannot be drawn.

MBW parameters cannot be calculated using the current software available at RBH. Although not ideal, raw data seems to be the best method for understanding simultaneous washout of both SF₆ and N₂.

9.2 Appendix 2

Brief explanation of Respiratory Mass Spectrometer (RMS) testing issues

The RMS was utilised in this thesis as it records both SF₆ and N₂ directly rather than calculating the gas concentration from other derived signals. Routine MBW testing on the RMS is completed with 4% SF₆. 1% SF₆ testing was also completed with both devices joined in series to identify if there were any equipment differences.

- The Inn/SF₆ was unable to record and realistically measure 4% SF₆. The Inn/SF₆ cannot be calibrated to 4% SF₆ as it is outside of the detection range for the gas analyser (the maximal range of the gas analyser is 1% SF₆). 4% SF₆ washout on both the RMS and Inn/SF₆ cannot be completed.
- The RMS is unable to detect 0.2% SF₆, even when changing the specifications on the RMS to be more sensitive to the molecular properties of SF₆. 0.2% SF₆ washout on both the RMS and Inn/SF₆ cannot be completed.

As the Inn/SF₆ can detect up to 1% and in theory so should the RMS (it has the detection limits to record 1% SF₆ but the signal to noise ratio may increase as the tracer gas decreases before 1% over the washout). The researcher therefore ordered a gas mix of 1% SF₆ to complete a dual gas MBW with all 3 devices in situ.

- The laptop would not connect to the RMS in order to run the MBW program where the RMS would detect flow and gas concentration. With advice from Robert Crawford, a RMS engineer as well as Emma Raywood, a respiratory physiologist from Great Ormond Street Hospital, the issue was not resolved as it involved substantive hardware changes, which would not be feasible within the time frame of this work. The raw gas concentration values, breath number and time to washout were utilised to determine gas differences.
- The raw data file (American Standard Code for Information Interchange (ASCII)) was unable to be exported from the RMS. The RMS Engineer concluded it may be because of missing or faulty hardware that transcribes a signal into quantifiable data, that would also enable connection to a laptop but this was unable to be replaced and or fix the error.