



Deposited via The University of Sheffield.

White Rose Research Online URL for this paper:

<https://eprints.whiterose.ac.uk/id/eprint/170740/>

Version: Accepted Version

---

**Article:**

Mallallah, F., Packham, A., Lee, E. et al. (2021) Is hyperpolarised gas magnetic resonance imaging a valid and reliable tool to detect lung health in cystic fibrosis patients? a cosmin systematic review. *Journal of Cystic Fibrosis*, 20 (6). pp. 906-919. ISSN: 1569-1993

<https://doi.org/10.1016/j.jcf.2020.12.020>

---

Article available under the terms of the CC-BY-NC-ND licence  
(<https://creativecommons.org/licenses/by-nc-nd/4.0/>).

**Reuse**

This article is distributed under the terms of the Creative Commons Attribution-NonCommercial-NoDerivs (CC BY-NC-ND) licence. This licence only allows you to download this work and share it with others as long as you credit the authors, but you can't change the article in any way or use it commercially. More information and the full terms of the licence here: <https://creativecommons.org/licenses/>

**Takedown**

If you consider content in White Rose Research Online to be in breach of UK law, please notify us by emailing [eprints@whiterose.ac.uk](mailto:eprints@whiterose.ac.uk) including the URL of the record and the reason for the withdrawal request.

1 **Is hyperpolarised gas magnetic resonance imaging a valid and**  
2 **reliable tool to detect lung health in cystic fibrosis patients? A**  
3 **COSMIN systematic review.**

4 Fatmah Mallallah<sup>1</sup>, Anna Packham<sup>2</sup>, Ellen Lee<sup>3</sup>, Daniel Hind<sup>3</sup>.

5

6 1. Al-Adan Hospital, Ahmadi Governorate, Kuwait.

7 2. Sheffield Children's NHS Foundation Trust, Sheffield, UK.

8 3. Sheffield Clinical Trials Research Unit, Sheffield, UK.

9

10 Email addresses:

11

12

13 e.lee@sheffield.ac.uk

14 d.hind@sheffield.ac.uk

15

---

16 **Abstract**

17

18 This paper systematically reviewed the literature reporting the validity and reliability of  
19 hyperpolarised gas MRI as a marker of lung health in cystic fibrosis (CF). MEDLINE,  
20 EMBASE and grey literature were searched for studies assessing the measurement  
21 properties of hyperpolarised helium-3 or xenon-129 MRI. The COSMIN risk of bias  
22 tool was used to critically appraise eligible studies. Findings show hyperpolarised gas  
23 MRI was able to detect structural and functional abnormalities in the lungs, detect  
24 response to treatments, and is more sensitive than FEV<sub>1</sub> in detecting ventilation  
25 defects in CF patients. There was moderately robust evidence for construct validity of  
26 hyperpolarised gas MRI, although evidence for other types of validity is currently low.  
27 Nonetheless, high quality studies concluded that hyperpolarised gas MRI is a reliable  
28 tool and test results are reproducible in CF patients. Hyperpolarised gas MRI is a  
29 promising tool for detecting early CF pulmonary disease and for longitudinal  
30 monitoring of CF.

31

32

33

34

## 35 Introduction

36

37 Cystic Fibrosis<sup>4</sup> (CF) involves deteriorations in lung health due to the inability of the  
38 airways to clear accumulating mucus, making lungs more prone to respiratory tract  
39 infection and sputum production[1]. Spirometry, plethysmography, and multiple breath  
40 nitrogen washout (MBNW) (which measures lung clearance index (LCI)), are used in  
41 routine care to assess the severity of the disease and measure changes in lung  
42 volume. Computerized Tomography (CT), radiography and Magnetic Resonance  
43 Imaging (MRI) of the thorax can examine changes in lung structure, but cannot be  
44 used routinely to monitor the progression of the disease for safety and cost  
45 reasons[2,3].

46 Hyperpolarised (HP) gas magnetic resonance imaging (MRI) provides detailed  
47 resolution images by visualizing the distribution of the HP gas after inhalation [3–6].  
48 Small areas of hypoventilation in the lungs give rise to a lower signal, quantified as  
49 ventilation defect percent (VDP). VDP can be quantified using different measurements

---

CF: Cystic Fibrosis, MBNW : Multiple Breath Nitrogen Washout , LCI: Lung Clearance Index, CT: Computerized Tomography, MRI: Magnetic Resonance Imaging, HP: Hyperpolarised, VDP: Ventilation Defect Percent, FEV<sub>1</sub>: Forced Expiratory Volume in 1 second, <sup>3</sup>He: Helium-3 gas, <sup>129</sup>Xe: Xenon-129, COSMIN: COnsensus-based Standards for the selection of health Measurement Instruments, PROMs: Patient-Reported Outcome Measures, ICC: Intra-Class coefficient, SDC: Smallest Detectable Change, LoA: Limits of Agreement, MIC: Minimal Important Change, GRADE: Grading of Recommendations Assessment, Development, and Evaluation, ANOVA: Analysis of Variance, SEM: Standard Error of Measurement, 95% CI: 95% Confidence Intervals, AUC: Area Under the Curve, PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-Analyses, RV/TLC: Residual Volume/Total Lung Capacity, RoB: Risk of Bias, CPT: Chest physiotherapy, ROC: Receiver operating characteristic.

50 such as k-means clustering, whole lung signal fraction, fuzzy c-means and linear  
51 binning[7]. VDP can be compared with other pulmonary function tests such as FEV<sub>1</sub>  
52 to get a better understanding of lung health[6]. Historically, the technique used helium-  
53 3 gas (<sup>3</sup>He), but its relatively high cost and low availability has led to the increased use  
54 of xenon-129 (<sup>129</sup>Xe). <sup>129</sup>Xe also dissolves more efficiently in the blood, providing better  
55 gas exchange information [6].

56

57 HP gas MRI has the potential to complement existing tests [4,5], but its current use is  
58 largely restricted to research purposes [8]. A systematic review of research on its  
59 measurement properties is needed to inform decisions about wider adoption. This  
60 paper aimed to systematically review primary research studies assessing the validity  
61 and reliability of HP <sup>129</sup>Xe or <sup>3</sup>He MRI as a marker of lung health.

62

## 63 **Methods (1350)**

64 The review was registered on PROSPERO database (CRD42019129588) before  
65 starting data extraction.

66

67

### 68 [Eligibility criteria](#)

69 Studies were eligible if they recruited people with CF aged 5 and over, irrespective of  
70 disease progression. Studies including patients with other conditions were included if

71 the data for the CF group could be disaggregated. Eligible studies assessed the  
72 reliability and validity of HP  $^3\text{He}$  or  $^{129}\text{Xe}$  MRI. To be eligible a study had to report a  
73 summary statistic pertaining to at least one of the following: internal consistency; test-  
74 retest reliability; measurement error; content validity; construct validity; criterion  
75 validity [9,10]. FEV<sub>1</sub> was used as a criterion measure of validity – this being the gold  
76 standard measure of pulmonary function in clinical practice [11]. All studies aiming to  
77 develop or assess the measurement properties of HP gas MRI were included. There  
78 was no restriction on publication type; conference abstracts and theses were included.  
79 Studies using animal models and studies only assessing the feasibility or tolerability  
80 of HP gas MRI were excluded. Studies using HP gas MRI to validate another measure,  
81 or as an outcome were also excluded. Only papers published in the English language  
82 were included, due to resource constraints.

83

#### 84 [Systematic Literature Search](#)

85 We searched MEDLINE and EMBASE via Ovid from inception to 21 August 2020, with  
86 no date restrictions, as well as EThOS for theses and Google Scholar for grey  
87 literature. We screened reference lists of eligible studies to identify further studies.  
88 Where necessary, we contacted authors to access unpublished data and identify  
89 further eligible studies. The combined thesaurus and free text terms related to the  
90 population and tests. The full electronic search strategy is on PROSPERO database  
91 (CRD42019129588). Two authors (FM, AP) independently screened titles and  
92 abstracts, then potentially relevant full-text articles for eligibility. Disagreements were  
93 resolved by a third reviewer (DH).

94

95 FM and AP extracted study characteristics (study objectives, design, sample size, age  
96 [however reported], and comparators) and summary statistics related to reliability,  
97 validity and responsiveness. For responsiveness, we assessed only ability of  
98 hyperpolarised gas MRI to detect changes in the lungs after treatment. We used the  
99 primary research studies' own hypotheses to assess construct validity, as no  
100 hypothesis had been set by the research team prior to data collection.

101

#### 102 [Risk of Bias Assessment](#)

103 FM and AP assessed risk of bias using the COSMIN risk of bias checklist [12] (Table  
104 1), with disagreements resolved by consensus. Although the COSMIN system was  
105 developed to assess the measurement properties of survey instruments  
106 (questionnaires), the underlying statistics used are the same as those used to evaluate  
107 the measurement properties of imaging techniques, and has sometimes been used  
108 for this purpose[13,14].

109

#### 110 [Rating the Evidence using COSMIN Criteria of Good Measurement Properties](#)

111 The COSMIN criteria of good measurement were used to rate study results as  
112 sufficient, indeterminate and insufficient evidence of reliability or validity [15,16]. For  
113 test-retest reliability, an intra-class correlation (ICC) of 0.7 was rated sufficient; studies  
114 presenting no ICC were rated indeterminate. Sufficient evidence of an adequate  
115 measurement error required the smallest detectable change (SDC) or limits of  
116 agreement (LoA) to be less than the minimal important change (MIC). In the absence  
117 of the MIC, the findings were deemed insufficient evidence of measurement error. For

118 criterion validity, correlation with the gold standard (FEV1) should be 0.7 or above to  
119 be rated sufficient. If this was not calculated, the study results would be rated  
120 indeterminate. For hypothesis testing (convergent validity) of construct validity, the  
121 results of the study should be in accordance with the study hypothesis to be rated  
122 sufficient. If no hypothesis was reported, the results would be rated indeterminate  
123 (Table 1).

#### 124 [Grading the Evidence using GRADE Approach](#)

125 The overall quality of evidence was graded as very low (very little confidence in  
126 measurement property estimate), low (limited confidence in measurement property  
127 estimate), moderate (moderately confidence in measurement property estimate) or  
128 high (very confident that the measurement property estimate is close to the true  
129 measurement property estimate) using the Grading of Recommendations  
130 Assessment, Development, and Evaluation (GRADE) system [15,16]. Risk of bias,  
131 inconsistency, imprecision, and indirectness were used to determine the grade of the  
132 quality of evidence. Each measurement property begins on the High level, and may  
133 then be downgraded levels to moderate, low or very low as appropriate.

134 Risk of bias was assessed as: 1) no risk of bias (multiple studies with adequate risk of  
135 bias/at least one of very good quality); 2) serious (multiple studies of doubtful quality/at  
136 least one study with adequate quality); 3) very serious (multiple studies with  
137 inadequate risk of bias/at least one study with of doubtful quality); or, 4) extremely  
138 serious (one study with inadequate quality) (Table 1) [15,16].

139 Inconsistency was assessed as 1) acceptable (>75% study results in accordance), 2)  
140 serious (<75% study results in accordance), or 3) very serious (if all studies' results  
141 were insufficient). Imprecision refers to the total sample size and was assessed as 1)  
142 acceptable (n>100), 2) serious (n= 50 to 100), or 3) very serious (n<50). Indirectness  
143 refers to the study population including participants from other populations than the  
144 one of interest, and was assessed as 1) acceptable (only CF participants) 2) serious  
145 (healthy controls included in sample) and 3) very serious (This is not applicable to this  
146 study as only studies which included CF patients in the sample were included) (Table  
147 1).

#### 148 [Summary Statistics Extracted](#)

149 We extracted and summarised summary statistics. Reliability was measured by  
150 intraclass correlation (ICC), Bland–Altman, analysis of variance (ANOVA) and  
151 measurement error [9]. Measurement error was measured by standard error of  
152 measurement (SEM), Smallest Detectable Change (SDC) or Limits of Agreement  
153 (LoA) [9,10]. For the different types of validity: construct validity was measured by  
154 spearman rank correlation [10]; criterion validity was measured by standard correlation  
155 such as Pearson correlation, and area under the curve by calculating the sensitivity  
156 and specificity of the instruments used [10]; Responsiveness was assessed using 95%  
157 confidence intervals (95% CI), P-values and mean difference before and after the  
158 given treatment. This was assessing whether the HP gas MRI detected any changes  
159 in the lungs in response to the treatment.

160



**Table 1: COSMIN Definitions and Methodology**

| Reliability/ validity measure | Definition   | RoB Checklist  | Good Measurement Properties   | Grading Quality of Evidence (GRADE Approach)   |
|-------------------------------|--|--|---|--|
| <b>Reliability</b>            |  |  |   |  |
| Measurement error             | The random and systematic error of a patient's result that is not associated to the true change in the construct to be measured. Measured by SEM, SDC or Limits of Agreement LoA [8,9]   | <ul style="list-style-type: none"> <li>· Patients stable in interim period</li> <li>· Time interval between scans appropriate</li> <li>· Test conditions similar for measurements</li> <li>· Measurement error: SEM, SDC of LoA calculated</li> <li>· Test retest: ICC calculated</li> </ul> | <ul style="list-style-type: none"> <li>+ SDC or LoA &lt; MIC<sup>5</sup></li> <li>? MIC not defined</li> <li>- SDC or LoA &gt; MIC<sup>5</sup></li> </ul>   | Number of levels to downgrade according to seriousness of each assessment:<br><br>Risk of bias<br>-0 Acceptable<br>-1 Serious<br>-2 Very serious<br>-3 Extremely serious |
| Test- retest                  | The reproducibility of results if the test is repeated over time. Measured by intraclass correlation (ICC), Bland-Altman, analysis of variance (ANOVA) and measurement error [9].  | <ul style="list-style-type: none"> <li>· Patients stable in interim period</li> <li>· Time interval between scans appropriate</li> <li>· Test conditions similar for measurements</li> <li>· Measurement error: SEM, SDC of LoA calculated</li> <li>· Test retest: ICC calculated</li> </ul> | <ul style="list-style-type: none"> <li>+ ICC or weighted Kappa ≥ 0.70</li> <li>? ICC or weighted Kappa not reported</li> <li>- ICC or weighted Kappa &lt; 0.70</li> </ul>   |  |
| <b>Validity</b>               |  |  |   |  |
| Construct (convergent)        | The degree to which the instrument relates to a measure it is hypothesised to have a strong relationship with. Measured by spearman rank correlation.  | <ul style="list-style-type: none"> <li>· Is it clear what the comparator (FEV<sub>1</sub>) measures</li> <li>· Were measurement properties of comparator (FEV<sub>1</sub>) sufficient</li> <li>· Design and statistical methods adequate for hypotheses to be tested</li> </ul>              | <ul style="list-style-type: none"> <li>+ ≥ 75% study results in accordance with the study hypothesis</li> <li>? No hypothesis defined by study authors</li> <li>- &lt; 75% study results in accordance with the study hypothesis</li> </ul> | Inconsistency<br>-0 Acceptable<br>-1 Serious<br>-2 Very serious  |
| Criterion                     | The extent to which the results of an instrument reflect the gold standard measurement (FEV <sub>1</sub> ). Measured by standard correlation such as Pearson correlation, and AUC by calculating sensitivity and specificity of instruments used [9] | <ul style="list-style-type: none"> <li>· Correlations or AUC calculated</li> </ul>   | <ul style="list-style-type: none"> <li>+ Correlation with FEV<sub>1</sub> ≥ 0.70 OR AUC ≥ 0.70</li> <li>? Not all information for '+' reported</li> <li>- Correlation with FEV<sub>1</sub> &lt; 0.70 OR AUC &lt; 0.70</li> </ul>            |  |
| Responsiveness                | The extent to which an instrument is able to detect a clinically important change in the concept being measured.   | <ul style="list-style-type: none"> <li>· Correlations between change scores or AUC calculated</li> </ul>   | <ul style="list-style-type: none"> <li>+ ≥ 75% study results in accordance with the hypothesis</li> <li>? No hypothesis defined (by the review team)</li> <li>- The result is not in accordance with the hypothesis</li> </ul>              | Imprecision<br>-0 Acceptable<br>-1 total n=50-100<br>-2 total n<50<br><br>Indirectness<br>-0 Acceptable<br>-1 Serious<br>-2 Very serious                                 |

162 Table Note: Information in table taken from COSMIN Manual for Systematic Review of PROMs [12][15][16].

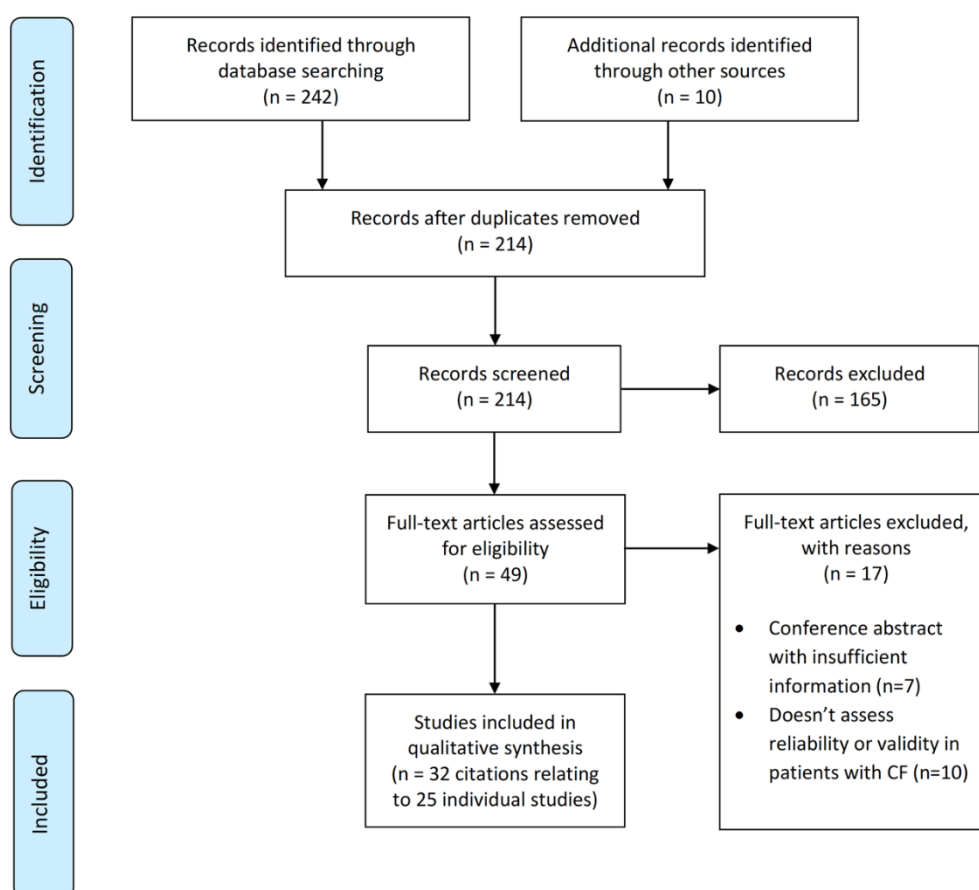
163 SEM = standard area of measurement, SDC = smallest detectable change, LoA = limits of agreement, AUC = area under receiver operator curve, ICC = intraclass correlation coefficient, MIC =

164 minimal important change. + = sufficient, ? = indeterminate, - = insufficient.

165

166 **Results**

167 Following the PRISMA reporting guidelines[17], after the elimination of duplicates, the  
 168 searches retrieved 204 articles through the electronic bibliographic databases and 10  
 169 citations through grey literature searching (total N=214) (Figure 1). After eligibility  
 170 screening, 49 full-text articles were retrieved for eligibility assessment. Of these, 32  
 171 articles, representing 25 unique studies met the eligibility criteria and were included in  
 172 the review (Table 2). Reasons for exclusion of studies at the full-text stage are given  
 173 in Appendix 1.



174

175

Figure 1: PRISMA 2009 Flow Diagram for Study Selection

**Table 2: Study Characteristics by Validity/ Reliability Measure**

| Author (Country)                       | <sup>3</sup> He or <sup>129</sup> Xe  | Study Design                           | Study Duration  | CF Sample Size    | CF Age Mean/Median (SD/Range)                           | Comparator Device   |
|--|---------------------------------------|--|---|-------------------|---|---|
| <b>Measurement error</b>               |                                       |  |   |                   |   |   |
| Kirby et al, 2011, Canada [18]         | <sup>3</sup> He                       | Case series                            | 2 scans in 1 week   | 12                | Mean = 26 (range 18 to 41)                              | Spirometry; plethysmography   |
| <b>Test-retest reliability</b>         |                                       |  |   |                   |   |   |
| Woodhouse et al, 2009, UK [19]         | <sup>3</sup> He                       | Cross-sectional                        | 2 scans in 1 session  | 5                 | Mean = 11 (range 6 to 15)                               | Spirometry  |
| Choy et al, 2010, Canada [20][21]      | <sup>3</sup> He                       | Pilot study                            | 2 scans in 1 week   | 8                 | Mean = 25 (SD=8)  | Spirometry  |
| Bannier et al, 2010, France [22]       | <sup>3</sup> He                       | Cross-sectional                        | 2 scans in 1 session  | 10                | Mean = 10.2 (range 8 to 16)                             | Spirometry (CPT was done for all patients to check changes in HP MRI) |
| O'Sullivan et al, 2014, Canada [23]    | <sup>3</sup> He                       | Case series                            | 4 scans in 4 weeks  | 5                 | NA  | Spirometry  |
| Zha et al, 2019, USA [8]               | <sup>3</sup> He                       | Cross-sectional and Case series study  | 2 scans in 2 weeks  | 7                 | Mean = 23.8 (SD=10.5)                                   | Spirometry  |
| Couch et al, 2019, Canada and USA [24] | <sup>129</sup> Xe                     | Retrospective analysis                 | 2 scans   | CF = 18<br>HC = 8 | CF Mean = 13.1 (SD=2.3)<br>Healthy Mean = 12.7 (SD=2.3) | Spirometry, plethysmography; MBNW                                     |
| Smith et al, 2020, UK [25]             | <sup>129</sup> Xe                     | Case series                            | 1 scan at baseline and at 16 month follow up (n=18)<br>2 scans at baseline and at 16 month follow up (n=11) | 29                | Mean = 23.0 (SD=11.1)                                   | spirometry, plethysmography; MBNW                                     |
| Smith et al, 2019, UK [26,27]          | <sup>3</sup> He and <sup>129</sup> Xe | Cross-sectional and Longitudinal study | 2 scans in 20 months  | 14                | Median 17.4 (range 6.4 – 47.5)                          | Spirometry; MBNW  |
| <b>Criterion validity</b>              |                                       |  |   |                   |   |   |
| Koumellis et al, 2005, USA [28]        | <sup>3</sup> He                       | Cross-sectional                        | 1 scan  | 8                 | Mean = 11.4 (range 6 to 15)                             | Spirometry  |

|                                   |                   |                 |  |                    |  |                                   |
|-----------------------------------|-------------------|-----------------|--|--------------------|--|-----------------------------------|
| Mentore et al, 2005, USA [29]     | <sup>3</sup> He   | Case series     | 3 scans for 3 different treatments         | CF= 15<br>HC = 16  | CF Mean = 21 (range 15 to 33)<br>Healthy Mean = 25 (range 21 to 33)                                      | Spirometry                        |
| Van Beek et al, 2006, UK [30]     | <sup>3</sup> He   | Cross-sectional | 1 scan                                     | 18                 | Median 12.1 (range 5 to 17)  | Spirometry; chest X-ray           |
| Woodhouse et al, 2009, UK [19]    | <sup>3</sup> He   | Cross-sectional | 2 scans in 1 session                       | 14                 | Reproducibility cohort:<br>Mean = 11 (range 6 to 15)<br>Intervention cohort:<br>Mean = 9 (range 5 to 15) | Spirometry                        |
| Choy et al, 2010, Canada [20][21] | <sup>3</sup> He   | Pilot study     | 2 scans in 1 week                          | 8                  | Mean 25 (SD=8)   | Spirometry                        |
| Kirby et al, 2013, Canada [31]    | <sup>3</sup> He   | Case series     | 2 scans in 1 week                          | 11                 | Mean 27 (SD=8)   | Spirometry; plethysmography       |
| Paulin et al, 2015, Canada [32]   | <sup>3</sup> He   | Case series     | 3 scans in 4 years                         | 5                  | 28 (range 20 to 36)  | Spirometry; plethysmography       |
| Hardy et al, 2016, UK [33]        | <sup>3</sup> He   | Cross-sectional | 1 scan                                     | CF= 18<br>HC = 30  | CF Mean = 14.7 (SD=2.3)<br>Healthy Mean = 14.6 (SD=1.4)  | Spirometry; plethysmography; MBNW |
| Marshall et al, 2017, UK [34]     | <sup>3</sup> He   | Cross-sectional | 1 scan                                     | CF = 19<br>HC = 10 | CF Mean = 10.9 (SD=2.5)<br>Healthy Mean= 11.3 (SD=2.8)   | Spirometry; plethysmography; MBNW |
| Smith et al, 2019, UK [35]        | <sup>3</sup> He   | Case series     | Scan at baseline and at 16 month follow up | 28                 | NA   | Spirometry; MBNW                  |
| Thomen et al, 2017, USA[36]       | <sup>129</sup> Xe | Cross-sectional | 1 scan                                     | CF = 11<br>HC = 11 | CF Mean = 12.5 (SD=2.3)<br>Healthy Mean = 11.5 (SD=3.2)  | Spirometry                        |
| Kanhere et al, 2017, Canada [37]  | <sup>129</sup> Xe | Cross-sectional | 1 scan                                     | CF = 10<br>HC = 5  | CF Mean = 13 (SD=2.5)<br>Healthy Mean = 12.4 (SD=2.4)  | Spirometry; plethysmography; MBNW |

|  |                                       |  |  |                            |  |   |
|--|---------------------------------------|--|--|----------------------------|--|---|
| Couch et al, 2019, Canada and USA [24] | <sup>129</sup> Xe                     | Retrospective analysis   | 1 scan   | CF = 18<br>HC = 8          | CF Mean = 13.1(SD=2.3)<br>Healthy Mean = 12.7 (SD=2.3) | spirometry, plethysmography; MBNW                                     |
| Rayment et al, 2019, Canada [38,39]    | <sup>129</sup> Xe                     | Cohort study   | 2 scans pre and post treatment over 3 weeks    | 15                         | Median of 14 (range 13.0 to 16.5)                      | Spirometry, plethysmography;MBNW                                      |
| <b>Construct validity</b>              |                                       |  |  |                            |  |   |
| McMahon et al, 2006, Ireland [4]       | <sup>3</sup> He                       | Case series  | 1 scan   | 8                          | Mean = 31.9 (range 20 to 46)                           | HRCT; spirometry  |
| Bannier et al, 2010, France [22]       | <sup>3</sup> He                       | Cross-sectional  | 2 scans in 1 session                           | 10                         | Mean = 10.2 (range 8 to 16)                            | Spirometry (CPT was done for all patients to check changes in HP MRI) |
| Kirby et al, 2011, Canada [18]         | <sup>3</sup> He                       | Case series  | 2 scans in 1 week                              | 12                         | Mean = 26 (range 18 to 41)                             | Spirometry; plethysmography   |
| Altes et al, 2012, USA [40–42]         | <sup>3</sup> He                       | Study A: Crossover clinical trial<br>Study B: open label trial | 5 scans in 48 weeks                            | Study A = 8<br>Study B = 9 | A Mean = 18.9<br>B Mean = 24.4                         | Spirometry; MBNW  |
| Smith et al, 2018, UK [43][44]         | <sup>3</sup> He                       | Cross-sectional  | 1 scan   | 32                         | Median 16.7 (range 6.4–43.1)                           | Spirometry; plethysmography; MBNW                                     |
| Smith et al, 2018, UK [45]             | <sup>3</sup> He                       | Case series  | 2 scans in 1.3-2 years                         | 14                         | Mean = 10.30 (SD=2.26)                                 | Spirometry; plethysmography; MBNW                                     |
| Zha et al, 2019, USA [8]               | <sup>3</sup> He                       | Cross-sectional and Case series study                          | 1 scan   | 17                         | Mean = 23.8 (SD=10.5)                                  | Spirometry  |
| Smith et al, 2019, UK [26,27]          | <sup>3</sup> He and <sup>129</sup> Xe | Cross-sectional and Longitudinal study                         | One (n=17) to two (n=14) scans in 20 months    | 31                         | Median 17.4 (range 6.4 – 47.5)                         | Spirometry; MBNW  |
| <b>Responsiveness</b>                  |                                       |  |  |                            |  |   |
| Bannier et al, 2010, France [22]       | <sup>3</sup> He                       | Cross-sectional  | 2 scans (pre- and post-CPT)                    | 10                         | Mean = 10.2 (range 8 to 16)                            | Spirometry (CPT was done for all patients to check changes in HP MRI) |
| Altes et al, 2012, USA [40–42]         | <sup>3</sup> He                       | Study A: Crossover clinical trial<br>Study B: open label trial | 5 scans in 48 weeks                            | Study A = 8<br>Study B = 9 | A Mean = 18.9<br>B Mean = 24.4                         | Spirometry; MBNW  |
| Rayment et al, 2019, Canada [38,39]    | <sup>129</sup> Xe                     | Cohort study   | 2 scans (pre- and post-treatment) over 3 weeks | 15                         | Median of 14 (range 13.0 to 16.5)                      | Spirometry, plethysmography; MBNW                                     |

|                                |                   |                 |                                       |                   |   |                                   |
|--------------------------------|-------------------|-----------------|---------------------------------------|-------------------|---|-----------------------------------|
| Smith et al,2020, UK [46]      | <sup>129</sup> Xe | Cross-sectional | 2 scans (pre- and post-exercise)      | 13                | Mean = 25 (SD=10)   | Spirometry, plethysmography; MBNW |
| Woodhouse et al, 2009, UK [19] | <sup>3</sup> He   | Cross-sectional | 2 scans (pre- and post-physiotherapy) | 9                 | Mean = 9 (range 5 to 15)  | Spirometry                        |
| Mentore et al, 2005, USA [29]  | <sup>3</sup> He   | Case series     | 3 scans for 3 different treatments    | CF= 15<br>HC = 16 | CF Mean = 21 (range 15 to 33)<br>Healthy Mean = 25 (range 21 to 33) | Spirometry                        |

177 Study Characteristics

178 The 25 included studies were published between 2005 and 2020. Six were conducted  
179 in the USA [8,24,28,29,36,40], seven in Canada [18,20,23,31,32,37,38], 10 in the UK  
180 [19,25,26,30,33–35,43,45,46], one in France [22], and one in Ireland [4]. Eighteen of  
181 the studies investigated HP <sup>3</sup>He MRI [4,8,18–20,22,23,28–35,40,43,45] six  
182 investigated HP <sup>129</sup>Xe MRI [24,25,36–38,46] and one investigated both HP <sup>3</sup>He and  
183 <sup>129</sup>Xe MRI [26] (Table 2).

184 There were 11 cross-sectional studies [8,19,22,25,28,30,33,34,36,37,43], six case  
185 series with a follow-up of less than twelve months [4,18,23,29,31] and four  
186 [25,32,35,45] with follow-up of between 14 months [25] and 4 years [32]. Two case  
187 series exposed participants to interventions such as nebulisers and chest  
188 physiotherapy prior to the MRI scan to understand treatment response [18,29]. One  
189 study presented a nested case series within a larger cross-sectional study [26]. There  
190 was one crossover clinical trial [40], There was one pilot study [20], one retrospective  
191 analysis study [24], and one cohort study [38].

192 Sample sizes ranged from five [23,32] to 32 [43] people with CF and from 5 [37] to 30  
193 [33] healthy individuals, in studies which used controls. The reported mean age of  
194 study populations ranged from 9 to 32. Reported median ages ranged from 12.1 to  
195 17.4.

196

197 [Test-retest Reliability](#)

198 Eight studies assessed test-retest reliability [8,19,20,22–26]; five using the ICC  
199 [8,20,22,24,25], four using Bland-Altman tests [8,19,25,26], and one using ANOVA  
200 [23] (Table 4). There was good evidence for the test-retest reliability of MRI in  
201 assessing VDP across three studies of very good [25] and adequate quality [20,22] in  
202 which the intraclass correlations were more than 0.9 (Table 4). The GRADE  
203 assessment was very low for studies using HP <sup>3</sup>He as there was inconsistency  
204 between studies, with <75% of the studies showing a strong correlation, and  
205 considerable imprecision (total n=49 study participants with cystic fibrosis) (Table 6).  
206 For studies using HP <sup>129</sup>Xe the GRADE assessment was low due to Risk of Bias (only  
207 two studies of doubtful and inadequate quality), imprecision (total n=61 study  
208 participants with cystic fibrosis) and inconsistency with <75% of the studies showing  
209 a strong correlation.

210

211 [Measurement Error](#)

212 There was poor evidence for measurement error from a single study of adequate  
213 quality [18] (Table 5). This study found the SDC for VDP (3%) to be higher than the  
214 MIC (2%), suggesting a considerable chance that the change detected by MRI for  
215 VDP was caused by measurement error [10,47]. The GRADE assessment was very  
216 low due to risk of bias (only one study of adequate quality), imprecision (total n=12  
217 study participants with cystic fibrosis) and difficulty in assessing consistency.

218

219 [Criterion Validity](#)

220 From the HP  $^3\text{He}$  and  $^{129}\text{Xe}$  MRI studies, fourteen assessed criterion validity  
221 [19,20,24,28–38] (Table 6). One study [28] presented p-values, which were not  
222 interpretable, rather than correlations. Evidence for criterion validity was mixed when  
223 using FEV<sub>1</sub> as a criterion of VDP. Four studies of very good quality showed a strong  
224 correlation ( $>0.7$ ) between FEV<sub>1</sub> and VDP [19,29,32,34], however six very good quality  
225 studies [20,24,31,33,36,38] found a weaker correlation ( $<0.7$ , range: 0.3-0.69). The  
226 GRADE assessment for criterion validity was low for studies using HP  $^3\text{He}$ , and very  
227 low for studies using HP  $^{129}\text{Xe}$ . These grades are due to inconsistency ( $<75\%$  showed  
228 strong correlation) and indirectness (total healthy controls:  $^{129}\text{Xe}$  n=24;  $^3\text{He}$  n=56),  
229 with HP  $^{129}\text{Xe}$  MRI being downgraded further for imprecision (total n=54 study  
230 participants with cystic fibrosis).

231 Of the three studies of very good quality which assessed LCI as a criterion for VDP,  
232 one study [24] showed a strong correlation ( $>0.7$ ) and two [33,38] a weaker correlation  
233 ( $<0.7$ , range: 0.13-0.61). Low correlations were found between HP gas MRI and body  
234 plethysmography; RV/TLC [34,38]; and CT scan score [34] (Table 6).

235

236 [Construct Validity](#)

237 Seven studies using HP  $^3\text{He}$  MRI assessed construct validity [4,8,18,22,40,43,45], in  
238 addition to one study which used both HP  $^3\text{He}$  and  $^{129}\text{Xe}$  [26] (Table 7). Four of the  
239 eight studies did not report a study hypothesis and were rated indeterminate  
240 [22,26,43,45] (Appendix 2). From the four studies which included a hypothesis, two  
241 studies of very good quality [4,40] found a strong correlation between FEV<sub>1</sub> and VDP  
242 ( $>0.7$ ) in accordance with their hypothesis, and one study of very good quality found

243 a weaker correlation ( $= -0.68$ ) [18]. The GRADE assessment for construct validity was  
244 moderate for the studies using HP  $^3\text{He}$ , after being downgraded for inconsistency, with  
245 less than 75% studies having sufficient results. The GRADE assessment for construct  
246 validity of studies using HP  $^{129}\text{Xe}$  was very low, due to difficulty assessing  
247 inconsistency of evidence from only one study, and imprecision (total  $n=31$  patients  
248 with cystic fibrosis).

249

250 From the studies of very good quality which assessed construct validity against other  
251 techniques, strong correlations ( $>0.7$ ) were found between hyperpolarised gas MRI  
252 and LCI in two studies [26,43], RV/TLC in one study [43], and CT scan in one study  
253 [4].

254

#### 255 [Responsiveness](#)

256 Six studies assessed responsiveness to treatment [19,22,29,38,40,46]. There were  
257 two studies [29,40] of very good quality which found HP  $^3\text{He}$  MRI was able to detect  
258 changes in ventilation volume and defects after treatment, however the evidence was  
259 rated as intermediate due to no hypothesis being set by the review team (Table 8).

260

261

262

263

264

265

266

267

268

269

270

271

272

273

274

275

**Table 4:** Test-Retest Reliability: Risk of Bias Within Studies, Good measurement properties according to COSMIN Checklist, Study Findings, and GRADE result

| Authors (Date)               | <sup>3</sup> He or <sup>129</sup> Xe | RoB        | Good Measurement Properties | Study Findings   | Number of patients in all of the studies | GRADE    |
|------------------------------|--------------------------------------|------------|-----------------------------|--|--|----------|
| Woodhouse et al (2009) [19]  | <sup>3</sup> He                      | Doubtful   | Indeterminate               | Bland–Altman analysis for both examinations, The mean difference between the two examinations = -0.037 (95% CI -7.7 to 0.15) |  |          |
| Choy et al (2010) [20]       | <sup>3</sup> He                      | Adequate   | Sufficient                  | ICC of Ventilation Gradients (VG3x3) of VDP = 0.92<br>ICC of Coefficients of Variation (CoV3x3) of VDP = 0.91                |  |          |
| Bannier et al (2010) [22]    | <sup>3</sup> He                      | Adequate   | Sufficient                  | ICC of VDP = 0.924   | (n=49)                                   | Very Low |
| O’Sullivan et al (2014) [23] | <sup>3</sup> He                      | Inadequate | Indeterminate               | ANOVA of VDP P= 0.2871   |  |          |
| Zha et al (2019) [8]         | <sup>3</sup> He                      | Doubtful   | Sufficient                  | Bland-Altman analysis of VDP = 0.023 (95% CI -0.06 to 0.105)<br>ICC of VDP = 0.95  |  |          |
| Smith et al, (2019) [26]     | <sup>3</sup> He                      | Doubtful   | Indeterminate               | Bland–Altman analysis of <sup>3</sup> He- <sup>129</sup> Xe HP MRI LoA = 8.9, -7.4%  |  |          |
| Couch et al (2019) [24]      | <sup>129</sup> Xe                    | Inadequate | Sufficient                  | ICC of VDP = 0.99  |  |          |
| Smith et al, (2020) [25]     | <sup>129</sup> Xe                    | Very good  | Sufficient                  | Bland-Altman analysis (LoA) of VDP = 0.8 [-7.0, 8.5]<br>ICC of VDP = 0.97 [0.94, 0.99]                                       | (n=61)                                   | Low      |

Smith et al, (2019) [26]      <sup>129</sup>Xe      Doubtful      Indeterminate      Bland–Altman analysis of <sup>3</sup>He-<sup>129</sup>Xe HP MRI LoA = 8.9, -7.4%

277

CI= Confidence intervals, ICC= Intraclass correlation coefficient, VDP= Ventilation defect percentage

**Table 5:** Measurement Error: Risk of Bias Within Studies, Good measurement properties according to COSMIN Checklist, Study Findings, and GRADE result

| Authors (Date)          | <sup>3</sup> He or <sup>129</sup> Xe | RoB      | Good Measurement Properties | Study Findings    | Number of patients in all of the studies | GRADE    |
|-------------------------|--------------------------------------|----------|-----------------------------|-------------------|--|----------|
| Kirby et al (2011) [18] | <sup>3</sup> He                      | Adequate | Insufficient                | SDC in VDP = 0.03 | (n=12)                                   | Very Low |

278

SDC= Smallest detectable change, VDP= Ventilation defect percentage

279

**Table 6:** Criterion Validity: Risk of Bias Within Studies, Good measurement properties according to COSMIN Checklist, Study Findings, and GRADE result

| Authors (Date)              | <sup>3</sup> He or <sup>129</sup> Xe | RoB        | Good Measurement Properties | Study Findings  | Number of patients in all of the studies | GRADE |
|-----------------------------|--------------------------------------|------------|-----------------------------|---|--|-------|
| Koumellis et al (2005) [28] | <sup>3</sup> He                      | Inadequate | Indeterminate               | Correlation between VDP and FEV1 not reported.<br>The six peripheral ROI measurements were averaged to obtain an index of flow in the peripheral lung, a good correlation with the FEV <sub>1</sub> analysed by means of a two-tailed Student's t-test was found P= 3.74 x 10 <sup>-5</sup> | (n=144)                                  | Low   |
| Mentore et al (2005) [29]   | <sup>3</sup> He                      | Very Good  | Sufficient                  | Correlation coefficients (r) for:<br>· VDP with FEV <sub>1</sub> = - 0.71   |  |       |
| Beek et al (2006) [30]      | <sup>3</sup> He                      | Doubtful   | Insufficient                | Correlation coefficients (r) for:<br>· <sup>3</sup> He MRI with FEV <sub>1</sub> = - 0.41   |  |       |

|                             |               |           |              |   |
|-----------------------------|---------------|-----------|--------------|---|
| Woodhouse et al (2009) [19] | $^3\text{He}$ | Very Good | Sufficient   | <p>Correlation coefficients (r) for:</p> <ul style="list-style-type: none"> <li>· <math>^3\text{He}</math> MRI with <math>\text{FEV}_1 = 0.98</math></li> <li>· <math>^3\text{He}</math> MRI with <math>\text{FEV}_1 = 0.82</math></li> </ul>   |
| Choy et al (2010) [20]      | $^3\text{He}$ | Very Good | Insufficient | <p>Correlation coefficients (r) for:</p> <ul style="list-style-type: none"> <li>· Ventilation Gradients (<math>\text{VG}_{3 \times 3}</math>) of VDP with <math>\text{FEV}_1 = 0.69</math></li> <li>· Coefficients of Variation (<math>\text{CoV}_{3 \times 3}</math>) of VDP with <math>\text{FEV}_1 = 0.66</math></li> </ul>            |
| Kirby et al (2013) [31]     | $^3\text{He}$ | Very Good | Sufficient   | <p>Correlation coefficients (r) for:</p> <ul style="list-style-type: none"> <li>· <math>\text{FEV}_1</math> with difference in whole lung apparent diffusion coefficient (ADC) (<math>^3\text{He}</math> MRI) = 0.67</li> <li>· <math>\text{FEV}_1</math> with previously ventilated ADC interior posterior difference = -0.75</li> </ul> |
| Paulin et al (2015) [32]    | $^3\text{He}$ | Very Good | Sufficient   | <p>Linear regression (<math>r^2</math>) for:</p> <ul style="list-style-type: none"> <li>· Baseline VDP with <math>\text{FEV}_1</math> after 4 years = 0.98</li> <li>· 4-year VDP with <math>\text{FEV}_1</math> after 4 years = 0.85</li> </ul>   |
| Hardy et al (2016) [33]     | $^3\text{He}$ | Very Good | Insufficient | <p>Correlation between VDP and <math>\text{FEV}_1</math> not reported.</p> <p>Correlation coefficients (r) for:</p> <ul style="list-style-type: none"> <li>· 13 ms ADC (VC W) with <math>\text{FEV}_1 = -0.39</math></li> </ul>   |

|                            |                   |            |               |  |        |          |
|----------------------------|-------------------|------------|---------------|--|--------|----------|
|                            |                   |            |               | <ul style="list-style-type: none"> <li>· 13 ms ADC (VC W) with LCI = -0.13</li> </ul>  |        |          |
| Marshall et al (2017) [34] | <sup>3</sup> He   | Inadequate | Sufficient    | <p>AUC for:</p> <ul style="list-style-type: none"> <li>· <sup>3</sup>He MRI VDP = 0.94</li> </ul> <p>Correlation coefficients (r) for:</p> <ul style="list-style-type: none"> <li>· VDP with RV/TLC = 0.61</li> <li>· VDP with CT gas trapping score = 0.58</li> <li>· VDP with LCI siting = 0.55</li> </ul> |        |          |
| Smith et al, (2019) [35]   | <sup>3</sup> He   | Inadequate | Indeterminate | <p>Correlation between VDP and FEV1 not reported</p> <p>Correlation coefficients (r) for:</p> <ul style="list-style-type: none"> <li>· The difference in VDP from baseline to the follow-up with the difference in LCI from baseline to the follow-up = 0.61</li> </ul>                                      |        |          |
| Kanhere et al (2017) [37]  | <sup>129</sup> Xe | Doubtful   | Insufficient  | <p>Coefficient of multiple correlation (<math>r^2</math>):</p> <ul style="list-style-type: none"> <li>· VDP with FEV<sub>1</sub> in all patients (CF and HC) = 0.31</li> <li>· VDP with LCI in all patients (CF and HC) = 0.88</li> </ul>  | (n=54) | Very Low |
| Thomen et al (2017) [36]   | <sup>129</sup> Xe | Very Good  | Insufficient  | <p>Correlation coefficients (r) for:</p> <ul style="list-style-type: none"> <li>· VDP with FEV<sub>1</sub> = - 0.54</li> </ul>   |        |          |
| Couch et al (2019) [24]    | <sup>129</sup> Xe | Very Good  | Insufficient  | <p>Linear regression (r<sup>2</sup>) for:</p>  |        |          |

|                           |                   |           |              |                             |  |
|---------------------------|-------------------|-----------|--------------|-----------------------------|--|
|                           |                   |           |              |                             | <ul style="list-style-type: none"> <li>· VDP with FEV<sub>1</sub> done by analyst 1 = 0.33</li> <li>· VDP with FEV<sub>1</sub> done by analyst 2 = 0.26</li> <li>· VDP with LCI done by analyst 1 = 0.76</li> <li>· VDP with LCI done by analyst 2 = 0.77</li> </ul> |
| Rayment et al (2019) [38] | <sup>129</sup> Xe | Very Good | Insufficient | Linear regression (r2) for: | <ul style="list-style-type: none"> <li>· VDP with FEV<sub>1</sub> = 0.30</li> <li>· VDP with LCI supine = 0.21</li> <li>· VDP with LCI seated = 0.38</li> <li>· VDP with RV/TLC = 0.34</li> </ul>  |

280  
281  
282

VDP= Ventilation defect percentage, FEV<sub>1</sub> = Forced expiratory volume in 1 second, ROI= Regions of interest, ADC = apparent diffusion coefficient, VCW= Weighted and volume corrected, RV= Residual volume, TLC= total lung capacity, AUC= Area under curve, CT= Chest tomography, LCI= lung clearance index, HC = healthy controls

**Table 7:** Construct Validity: Risk of Bias Within Studies, Good measurement properties according to COSMIN Checklist, Study Findings, and GRADE result

| Authors (Date)            | <sup>3</sup> He or <sup>129</sup> Xe | RoB       | Good Measurement Properties | Study Findings   | Number of patients in all of the studies | GRADE    |
|---------------------------|--------------------------------------|-----------|-----------------------------|--|--|----------|
| McMahon et al (2006) [4]  | <sup>3</sup> He                      | Very Good | Sufficient                  | Spearman rank correlation (ρ) for: <ul style="list-style-type: none"> <li>· FEV<sub>1</sub> with VDP = 0.86</li> <li>· HRCT with VDP = ± 0.89</li> </ul> | (n=148)                                  | Moderate |
| Bannier et al (2010) [22] | <sup>3</sup> He                      | Very Good | Insufficient                | Spearman rank correlation (ρ) for: <ul style="list-style-type: none"> <li>· FEV<sub>1</sub> with VDP = - 0.041</li> </ul>                                |  |          |
| Kirby et al (2011) [18]   | <sup>3</sup> He                      | Very Good | Insufficient                | Spearman rank correlation (ρ) for:   |  |          |

|                          |                   |           |               |   |        |          |
|--------------------------|-------------------|-----------|---------------|---|--------|----------|
| Altes et al (2012) [40]  | <sup>3</sup> He   | Very Good | Sufficient    | <ul style="list-style-type: none"> <li>· FEV<sub>1</sub> with VDP = - 0.68</li> </ul> <p>Spearman rank correlation (ρ) for:</p> <ul style="list-style-type: none"> <li>· FEV<sub>1</sub> with VDP = - 0.85</li> <li>· Part A of the study, Spearman rank correlation (ρ) for:                             <ul style="list-style-type: none"> <li>▪ FEV<sub>1</sub> with VDP = - 0.52</li> </ul> </li> <li>· Part B of the study, Spearman rank correlation (ρ) for:                             <ul style="list-style-type: none"> <li>• FEV<sub>1</sub> with VDP = - 0.67</li> </ul> </li> </ul> |        |          |
| Smith et al (2018) [43]  | <sup>3</sup> He   | Very Good | Indeterminate | <p>Spearman rank correlation (ρ) for:</p> <ul style="list-style-type: none"> <li>· VDP with FEV<sub>1</sub> = - 0.79</li> <li>· VDP with LCI = 0.89</li> <li>· VDP with RV/TLC = 0.80</li> </ul>  |        |          |
| Smith et al (2018) [45]  | <sup>3</sup> He   | Doubtful  | Indeterminate | <p>Correlation between VDP and FEV<sub>1</sub> not reported</p> <p>Spearman rank correlation (ρ) for:</p> <ul style="list-style-type: none"> <li>· VDP and LCI at baseline = 0.66</li> <li>· VDP and LCI at visit 2 = 0.82</li> <li>· The percentage change in VDP from baseline to visit 2 = 0.60</li> </ul>   |        |          |
| Zha et al (2019) [8]     | <sup>3</sup> He   | Doubtful  | Sufficient    | <p>Spearman rank correlation (ρ) for:</p> <ul style="list-style-type: none"> <li>· VDP with FEV<sub>1</sub> = - 0.75</li> </ul>   |        |          |
| Smith et al, (2019) [26] | <sup>3</sup> He   | Very Good | Indeterminate | <p>Spearman rank correlation (ρ) for:</p> <ul style="list-style-type: none"> <li>· VDP with FEV<sub>1</sub> = - 0.78</li> <li>· VDP with LCI = 0.88</li> </ul>  |        |          |
| Smith et al, (2019) [26] | <sup>129</sup> Xe | Very Good | Indeterminate | <p>Spearman rank correlation (ρ) for:</p>   | (n=31) | Very Low |

- VDP with FEV<sub>1</sub> = - 0.79
- VDP with LCI = 0.88

283

FEV<sub>1</sub> = Forced expiratory volume in 1 second, VDP= Ventilation defect percentage, LCI= Lung clearance index, RV= Residual volume, TLC= total lung capacity

**Table 8:** Responsiveness: Risk of Bias Within Studies, Good measurement properties according to COSMIN Checklist, Study Findings, and GRADE result

| Authors (Date)              | <sup>3</sup> He or <sup>129</sup> Xe | RoB        | Good Measurement Properties | GRADE  |     |
|-----------------------------|--------------------------------------|------------|-----------------------------|--|-----|
| Mentore et al (2005) [29]   | <sup>3</sup> He                      | Very Good  | Indeterminate               | Correlation between change in VDP and FEV <sub>1</sub> (% predicted) from baseline to after albuterol, DNase, and chest physical therapy was weak (r = -0.13).   |     |
| Woodhouse et al (2009) [19] | <sup>3</sup> He                      | Inadequate | Indeterminate               | The was no significant changes in total ventilation volume before and after CPT that was detected using hyperpolarised helium MRI (P value = 0.36)   | N/A |
| Bannier et al (2010) [22]   | <sup>3</sup> He                      | Inadequate | Indeterminate               | VDP before and after CPT did not change (P value > 0.10);<br><ul style="list-style-type: none"> <li>· VDP mean before CPT = 5.1 (1.9)</li> <li>· VDP mean after CPT = 5.1 (1.1)</li> </ul>   | N/A |
| Altes et al (2012) [40]     | <sup>3</sup> He                      | Very Good  | Indeterminate               | Part A: VDP was reduced by 8.2% from baseline (day 15) to after ivacaftor treatment (day 43), P value = 0.055 (r= -0.5238)   |     |
| Rayment et al (2019) [38]   | <sup>129</sup> Xe                    | Inadequate | Indeterminate               | The absolute mean change in VDP pre- post treatment was -3.0 (-5.0, -1.0)<br>The relative change (%) in VDP pre- post treatment was -44.2 (-60.2, -28.3)   |     |
| Smith et al, (2020) [46]    | <sup>129</sup> Xe                    | Inadequate | Indeterminate               | There was a small but significant reduction in the VDP (p = 0.04) after CPET when compared to baseline.<br><ul style="list-style-type: none"> <li>· VDP % before CPET = 7.3 [2.3, 25.8]</li> <li>· VDP % after CPET = 7.1 [2.4, 24.8]</li> </ul> | N/A |

284 VDP= Ventilation defect percentage, FEV<sub>1</sub> = Forced expiratory volume in 1 second, CPT= Chest physiotherapy, CPET= Cardiopulmonary exercise testing

285

286

287

288

289

290

291

292

293

294

295 [Summary of Risk of Bias Issues](#)

296 The greatest risk of bias for tests of reliability resulted from inadequate reporting of  
297 patient stability on the day of scans [8,18,19,21,22,26] and consistency in test  
298 conditions for all participants [8,18,19,26]. Reporting of the summary statistics  
299 necessary to understand measurement properties was often inadequate. For instance,  
300 ICCs, mean differences and 95% CIs were often not reported for correlations  
301 [23,24,30,37]. Risk of bias arose in studies of construct and criterion validity which did  
302 not correlate HP gas MRI against FEV<sub>1</sub> (the current gold standard) [8,28,34,35,45].  
303 Further to this, studies of responsiveness frequently failed to correlate changes in HP  
304 gas MRI with those observed in FEV<sub>1</sub> [19,22,38,46].

305

306 [Summary of Issues arising during Grading](#)

307 Evidence for reliability was frequently downgraded for inconsistency in study results  
308 and imprecision due to the studies' small sample size. Evidence for validity was  
309 downgraded for inconsistency in findings, indirectness due to heterogeneity in  
310 samples including healthy participants, and imprecision arising from low statistical  
311 power.

312 **Discussion**

313 This review found moderately robust evidence for construct validity of HP gas MRI as  
314 a marker of lung health in people with cystic fibrosis. Evidence for other types of  
315 validity and reliability is currently low. Nonetheless, high quality studies [4,20,22,27,43]  
316 concluded that HP gas MRI was a useful tool to detect lung ventilation defects, was

317 useful in tracing the functional and structural progression of cystic fibrosis, and test  
318 results were reproducible in cystic fibrosis patients.

319

320 HP gas MRI was able to detect ventilation defects in patients with normal FEV<sub>1</sub> results  
321 [22,24,31,34,36] and is better able to discriminate CF patients from healthy controls  
322 than FEV<sub>1</sub> [36], especially in children where the disease is still developing. While FEV<sub>1</sub>  
323 was sensitive in detecting obstruction in large airways, FEV<sub>1</sub> cannot detect ventilation  
324 defects in small airways [48,49]. That HP gas MRI can detect changes in VDP over  
325 short periods of time, indicates its potential in the management of CF [18,31].

326

327 It is important to note that there is currently a lack of standardisation in the acquisition  
328 and analysis of HP gas MRI. This is a limitation of this review, as the differences in the  
329 generation of VDP could mean the measures of validity and reliability are not be  
330 directly comparable[50][51].

331

332 Future research needs to serve both a policy-making and clinical audience, including  
333 those who still see FEV<sub>1</sub> as the gold standard and those who see it as an insensitive  
334 measure of lung health. Advocates of HP gas MRI have made the case, successfully,  
335 that it detects functional defects in CF patients better than other methods. To bring  
336 about a shift in clinical and policy norms requires an argument about why that matters  
337 in clinical terms and is cost-effective for health systems. In particular, overuse of  
338 imaging has adverse economic consequences and is burdensome for patients in  
339 terms of repeat tests and exposure to x-rays [52]. Qualitative research is needed to  
340 assess the degree to which cystic fibrosis specialists (respiratory physicians and

341 physiotherapists) consider HP gas MRI an adequate reflection of lung health (content  
342 validity). Given the long natural history of cystic fibrosis, decision-makers may require  
343 that changes in gas MRI-assessed lung health are validated against FEV<sub>1</sub> and other  
344 instruments, such as LCI, over a period of at least five years to demonstrate MRI's  
345 prognostic validity and clinical potential.

346

347 To reduce risk of bias, future studies must document patient stability on the day of  
348 testing, and report consistency in test conditions, for all participants. Authors of studies  
349 which reported patient stability defined this as no changes in respiratory symptoms or  
350 medications in the period leading up to participation in the study, which ranged from 1  
351 week to 4 weeks [20,23–25]. To improve the statistical quality of studies, when  
352 analysing correlations, the ICC, mean difference and 95% CI should be reported.

## 353 Conclusion

354 HP gas MRI is a promising tool for detecting early CF pulmonary disease and for  
355 longitudinal monitoring of the progression of the disease. It is more sensitive than  
356 FEV<sub>1</sub>, in detecting functional and structural ventilation defects in CF patients and is  
357 responsive to CF pulmonary treatments. Further validation is required against a range  
358 of measures in long-term studies to assess its prognostic value and cost-  
359 effectiveness.

360

## 361 Acknowledgements.

362 No specific funding supported this systematic review. DH conceived of the study. DH  
363 and FM designed the searches and, with AP, drafted the manuscript. DH and FM

364 designed the study. FM and AP ran searches, screened studies for eligibility, extracted  
365 data and critically appraised primary research studies. EL was the study statistician  
366 and contributed to the first and subsequent drafts of the manuscript. DH, AP, FM and  
367 EL commented on and approved the final manuscript.

368

369 **Conflicts of Interests.**

370 All authors declare that they have no competing interests.

371

372

373

374

375

376

377

378

379

380

381

382

383

384

385

386 **Reference List**

387

388 [1] Davies JC, Alton EFWF, Bush A. Cystic fibrosis. *BMJ* 2007;335:1255–9.  
389 <https://doi.org/10.1136/bmj.39391.713229.AD>.

390 [2] Mayo JR, Aldrich J, Müller NL. Radiation Exposure at Chest CT: A Statement  
391 of the Fleischner Society. *Radiology* 2003;228:15–21.  
392 <https://doi.org/10.1148/radiol.2281020874>.

393 [3] Donnelly LF, MacFall JR, McAdams HP, Majure JM, Smith J, Frush DP, et al.  
394 Cystic Fibrosis: Combined Hyperpolarized <sup>3</sup>He-enhanced and Conventional  
395 Proton MR Imaging in the Lung—Preliminary Observations. *Radiology*  
396 1999;212:885–9. <https://doi.org/10.1148/radiology.212.3.r99se20885>.

397 [4] McMahon CJ, Dodd JD, Hill C, Woodhouse N, Wild JM, Fischele S, et al.  
398 Hyperpolarized <sup>3</sup>helium magnetic resonance ventilation imaging of the lung in  
399 cystic fibrosis: comparison with high resolution CT and spirometry. *Eur Radiol*  
400 2006;16:2483–90. <https://doi.org/10.1007/s00330-006-0311-5>.

401 [5] Evans A, McCormack DG, Santyr G, Parraga G. Mapping and quantifying  
402 hyperpolarized <sup>3</sup>He magnetic resonance imaging apparent diffusion  
403 coefficient gradients. *J Appl Physiol* 2008;105:693–9.  
404 <https://doi.org/10.1152/jappphysiol.00178.2008>.

405 [6] Santyr G, Kanhere N, Morgado F, Rayment JH, Ratjen F, Couch MJ.  
406 Hyperpolarized Gas Magnetic Resonance Imaging of Pediatric Cystic Fibrosis  
407 Lung Disease. *Acad Radiol* 2019;26:344–54.  
408 <https://doi.org/10.1016/j.acra.2018.04.024>.

409 [7] Woodhouse N, Wild JM, Paley MNJ, Fischele S, Said Z, Swift AJ, et al.

- 410 Combined helium-3/proton magnetic resonance imaging measurement of  
411 ventilated lung volumes in smokers compared to never-smokers. *J Magn*  
412 *Reson Imaging* 2005;21:365–9. <https://doi.org/10.1002/jmri.20290>.
- 413 [8] Zha W, Nagle SK, Cadman R V., Schiebler ML, Fain SB. Three-dimensional  
414 Isotropic Functional Imaging of Cystic Fibrosis Using Oxygen-enhanced MRI:  
415 Comparison with Hyperpolarized 3 He MRI. *Radiology* 2019;290:229–37.  
416 <https://doi.org/10.1148/radiol.2018181148>.
- 417 [9] Bruton A, Conway JH, Holgate ST. Reliability: What is it, and how is it  
418 measured? *Physiotherapy* 2000;86:94–9. [https://doi.org/10.1016/S0031-](https://doi.org/10.1016/S0031-9406(05)61211-4)  
419 [9406\(05\)61211-4](https://doi.org/10.1016/S0031-9406(05)61211-4).
- 420 [10] Scholtes VA, Terwee CB, Poolman RW. What makes a measurement  
421 instrument valid and reliable? *Injury* 2011;42:236–40.  
422 <https://doi.org/10.1016/j.injury.2010.11.042>.
- 423 [11] Barreiro TJ, Perillo I. An approach to interpreting spirometry. *Am Fam*  
424 *Physician* 2004;69:1107–14.
- 425 [12] Mokkink LB, de Vet HCW, Prinsen CAC, Patrick DL, Alonso J, Bouter LM, et  
426 al. COSMIN Risk of Bias checklist for systematic reviews of Patient-Reported  
427 Outcome Measures. *Qual Life Res* 2018;27:1171–9.  
428 <https://doi.org/10.1007/s11136-017-1765-4>.
- 429 [13] Saltzherr MS, Selles RW, Bierma-Zeinstra SMA, Muradin GSR, Coert JH, van  
430 Neck JW, et al. Metric properties of advanced imaging methods in  
431 osteoarthritis of the hand: a systematic review. *Ann Rheum Dis* 2014;73:365–  
432 75. <https://doi.org/10.1136/annrheumdis-2012-202515>.
- 433 [14] Jaspers MEH, van Haasterecht L, van Zuijlen PPM, Mokkink LB. A systematic

- 434 review on the quality of measurement techniques for the assessment of burn  
435 wound depth or healing potential. *Burns* 2019;45:261–81.  
436 <https://doi.org/10.1016/j.burns.2018.05.015>.
- 437 [15] Terwee CB, Prinsen CAC, Chiarotto A, Westerman MJ, Patrick DL, Alonso J,  
438 et al. COSMIN methodology for evaluating the content validity of patient-  
439 reported outcome measures: a Delphi study. *Qual Life Res* 2018;27:1159–70.  
440 <https://doi.org/10.1007/s11136-018-1829-0>.
- 441 [16] Prinsen CAC, Mokkink LB, Bouter LM, Alonso J, Patrick DL, de Vet HCW, et  
442 al. COSMIN guideline for systematic reviews of patient-reported outcome  
443 measures. *Qual Life Res* 2018;27:1147–57. [https://doi.org/10.1007/s11136-](https://doi.org/10.1007/s11136-018-1798-3)  
444 [018-1798-3](https://doi.org/10.1007/s11136-018-1798-3).
- 445 [17] Moher D, Liberati A, Tetzlaff J, Altman DG. Preferred Reporting Items for  
446 Systematic Reviews and Meta-Analyses: The PRISMA Statement. *PLoS Med*  
447 2009;6:e1000097. <https://doi.org/10.1371/journal.pmed.1000097>.
- 448 [18] Kirby M, Svenningsen S, Ahmed H, Wheatley A, Etemad-Rezai R, Paterson  
449 NAM, et al. Quantitative Evaluation of Hyperpolarized Helium-3 Magnetic  
450 Resonance Imaging of Lung Function Variability in Cystic Fibrosis. *Acad*  
451 *Radiol* 2011;18:1006–13. <https://doi.org/10.1016/j.acra.2011.03.005>.
- 452 [19] Woodhouse N, Wild JM, van Beek EJR, Hoggard N, Barker N, Taylor CJ.  
453 Assessment of hyperpolarized <sup>3</sup>He lung MRI for regional evaluation of  
454 interventional therapy: A pilot study in pediatric cystic fibrosis. *J Magn Reson*  
455 *Imaging* 2009;30:981–8. <https://doi.org/10.1002/jmri.21949>.
- 456 [20] Choy S, Ahmed H, Wheatley A, McCormack DG, Parraga G. Development of  
457 spatial-temporal ventilation heterogeneity and probability analysis tools for

- 458 hyperpolarized 3 He magnetic resonance imaging. In: Molthen RC, Weaver  
459 JB, editors., 2010, p. 762613. <https://doi.org/10.1117/12.844587>.
- 460 [21] Ahmed H, Choy S, Wheatley A, Etemad-Rezai R, Paterson N, Parraga G.  
461 Mapping And Quantifying Temporal Dynamics Of Ventilation Heterogeneity In  
462 Cystic Fibrosis Using Hyperpolarized Helium-3 Magnetic Resonance Imaging:  
463 Developing New Imaging Measurements For Clinical Trials. B70. FINE Anal.  
464 Alveolar/airw. Struct. Funct., American Thoracic Society; 2010, p. A3642–  
465 A3642. [https://doi.org/10.1164/ajrccm-](https://doi.org/10.1164/ajrccm-conference.2010.181.1_MeetingAbstracts.A3642)  
466 [conference.2010.181.1\\_MeetingAbstracts.A3642](https://doi.org/10.1164/ajrccm-conference.2010.181.1_MeetingAbstracts.A3642).
- 467 [22] Bannier E, Cieslar K, Mosbah K, Aubert F, Duboeuf F, Salhi Z, et al.  
468 Hyperpolarized 3 He MR for Sensitive Imaging of Ventilation Function and  
469 Treatment Efficiency in Young Cystic Fibrosis Patients with Normal Lung  
470 Function. *Radiology* 2010;255:225–32.  
471 <https://doi.org/10.1148/radiol.09090039>.
- 472 [23] O’Sullivan B, Couch M, Roche JP, Walvick R, Zheng S, Baker D, et al.  
473 Assessment of Repeatability of Hyperpolarized Gas MR Ventilation Functional  
474 Imaging in Cystic Fibrosis. *Acad Radiol* 2014;21:1524–9.  
475 <https://doi.org/10.1016/j.acra.2014.07.008>.
- 476 [24] Couch MJ, Thomen R, Kanhere N, Hu R, Ratjen F, Woods J, et al. A two-  
477 center analysis of hyperpolarized 129Xe lung MRI in stable pediatric cystic  
478 fibrosis: Potential as a biomarker for multi-site trials. *J Cyst Fibros*  
479 2019;18:728–33. <https://doi.org/10.1016/j.jcf.2019.03.005>.
- 480 [25] Smith LJ, Horsley A, Bray J, Hughes PJC, Biancardi A, Norquay G, et al. The  
481 assessment of short and long term changes in lung function in CF using 129

- 482 Xe MRI. *Eur Respir J* 2020;2000441. <https://doi.org/10.1183/13993003.00441->  
483 2020.
- 484 [26] Smith L, Collier G, Marshall H, Hughes P, Biancardi A, Norquay G, et al. P213  
485 A comparison of ventilation MRI using hyperpolarised <sup>3</sup>He and <sup>129</sup>Xe to  
486 assess cystic fibrosis lung disease. *J Cyst Fibros* 2019;18:S117.  
487 [https://doi.org/10.1016/S1569-1993\(19\)30506-5](https://doi.org/10.1016/S1569-1993(19)30506-5).
- 488 [27] Smith L, Collier G, Marshall H, Hughes P, Biancardi A, Norquay G.  
489 Comparison of hyperpolarised <sup>3</sup>He and <sup>129</sup>Xe ventilation MRI to assess lung  
490 disease in cystic fibrosis. *Int. Soc. Magn. Reson. Med., International Society*  
491 *for Magnetic Resonance in Medicine*; 2019.
- 492 [28] Koumellis P, van Beek EJR, Woodhouse N, Fichele S, Swift AJ, Paley MNJ, et  
493 al. Quantitative analysis of regional airways obstruction using dynamic  
494 hyperpolarized<sup>3</sup>He MRI—Preliminary results in children with cystic fibrosis. *J*  
495 *Magn Reson Imaging* 2005;22:420–6. <https://doi.org/10.1002/jmri.20402>.
- 496 [29] Mentore K, Froh DK, de Lange EE, Brookeman JR, Paget-Brown AO, Altes  
497 TA. Hyperpolarized HHe <sup>3</sup> MRI of the Lung in Cystic Fibrosis. *Acad Radiol*  
498 2005;12:1423–9. <https://doi.org/10.1016/j.acra.2005.07.008>.
- 499 [30] van Beek EJR, Hill C, Woodhouse N, Fichele S, Fleming S, Howe B, et al.  
500 Assessment of lung disease in children with cystic fibrosis using  
501 hyperpolarized <sup>3</sup>-Helium MRI: comparison with Shwachman score, Chrispin-  
502 Norman score and spirometry. *Eur Radiol* 2007;17:1018–24.  
503 <https://doi.org/10.1007/s00330-006-0392-1>.
- 504 [31] Kirby M, Villemaire L, Ahmed H, Paterson NA, McCormack DG, Lewis JF.  
505 Diffusion-Weighted Hyperpolarized Helium-3 Magnetic Resonance Imaging In

- 506 Adult Cystic Fibrosis. *Am J Respir Crit Care Med* 2013;194:A2072.
- 507 [32] Paulin GA, Svenningsen S, Jobse BN, Mohan S, Kirby M, Lewis JF, et al.  
508 Differences in hyperpolarized <sup>3</sup>He ventilation imaging after 4 years in adults  
509 with cystic fibrosis. *J Magn Reson Imaging* 2015;41:1701–7.  
510 <https://doi.org/10.1002/jmri.24744>.
- 511 [33] Hardy SM. Study of hyperpolarised <sup>3</sup>He MRI diffusion on asthma and cystic  
512 fibrosis, and development of hyperpolarised <sup>129</sup>Xe MRI lung imaging.  
513 University of Nottingham, 2016.
- 514 [34] Marshall H, Horsley A, Taylor CJ, Smith L, Hughes D, Horn FC, et al.  
515 Detection of early subclinical lung disease in children with cystic fibrosis by  
516 lung ventilation imaging with hyperpolarised gas MRI. *Thorax* 2017;72:760–2.  
517 <https://doi.org/10.1136/thoraxjnl-2016-208948>.
- 518 [35] Smith L, Collier G, Marshall H, Hughes P, Biancardi A, Norquay G, et al. P212  
519 Ventilation MRI tracks longitudinal lung function changes in patients with cystic  
520 fibrosis and clinically stable FEV1 and Lung Clearance Index. *J Cyst Fibros*  
521 2019;18:S117. [https://doi.org/10.1016/S1569-1993\(19\)30505-3](https://doi.org/10.1016/S1569-1993(19)30505-3).
- 522 [36] Thomen RP, Walkup LL, Roach DJ, Cleveland ZI, Clancy JP, Woods JC.  
523 Hyperpolarized <sup>129</sup>Xe for investigation of mild cystic fibrosis lung disease in  
524 pediatric patients. *J Cyst Fibros* 2017;16:275–82.  
525 <https://doi.org/10.1016/j.jcf.2016.07.008>.
- 526 [37] Kanhere N, Couch MJ, Kowalik K, Zanette B, Rayment JH, Manson D, et al.  
527 Correlation of Lung Clearance Index with Hyperpolarized <sup>129</sup>Xe Magnetic  
528 Resonance Imaging in Pediatric Subjects with Cystic Fibrosis. *Am J Respir Crit*  
529 *Care Med* 2017;196:1073–5. <https://doi.org/10.1164/rccm.201611-2228LE>.

- 530 [38] Rayment JH, Couch MJ, McDonald N, Kanhere N, Manson D, Santyr G, et al.  
531 Hyperpolarised <sup>129</sup>Xe magnetic resonance imaging to monitor treatment  
532 response in children with cystic fibrosis. *Eur Respir J* 2019;53:1802188.  
533 <https://doi.org/10.1183/13993003.02188-2018>.
- 534 [39] Rayment JH, McDonald N, Kanhere N, Couch M, Santyr G, Ratjen FA.  
535 Posture-Dependence of the Lung Clearance Index Varies with Disease State  
536 in Cystic Fibrosis. A72. WHAT'S NEW Cyst. FIBROSIS, BPD OTHER  
537 Congenit. Pediatr. LUNG Dis., San Diego: American Thoracic Society; 2018, p.  
538 A2308–A2308.
- 539 [40] Altes T, Johnson MA, Miller GW, Mugler JP, Flors L, Mata J, et al. 46  
540 Hyperpolarized Gas MRI of ivacaftor therapy in subjects with cystic fibrosis  
541 who have the G551D-CFTR mutation. *J Cyst Fibros* 2012;11:S67.  
542 [https://doi.org/10.1016/S1569-1993\(12\)60215-X](https://doi.org/10.1016/S1569-1993(12)60215-X).
- 543 [41] Altes TA, Johnson M, Fidler M, Botfield M, Tustison NJ, Leiva-Salinas C, et al.  
544 Use of hyperpolarized helium-3 MRI to assess response to ivacaftor treatment  
545 in patients with cystic fibrosis. *J Cyst Fibros* 2017;16:267–74.  
546 <https://doi.org/10.1016/j.jcf.2016.12.004>.
- 547 [42] Altes T, Johnson M, Mugler III J, Miller GW, Flors L, Mata J, et al. The Effect  
548 Of Ivacaftor, An Investigational CFTR Potentiator, On Hyperpolarized Noble  
549 Gas Magnetic Resonance Imaging In Subjects With Cystic Fibrosis Who Have  
550 The G551D-CFTR Mutation. B35. Pathog. Clin. ISSUES Cyst. Fibros.,  
551 American Thoracic Society; 2012, p. A2814–A2814.  
552 [https://doi.org/10.1164/ajrccm-](https://doi.org/10.1164/ajrccm-conference.2012.185.1_MeetingAbstracts.A2814)  
553 [conference.2012.185.1\\_MeetingAbstracts.A2814](https://doi.org/10.1164/ajrccm-conference.2012.185.1_MeetingAbstracts.A2814).

- 554 [43] Smith LJ, Collier GJ, Marshall H, Hughes PJC, Biancardi AM, Wildman M, et  
555 al. Patterns of regional lung physiology in cystic fibrosis using ventilation  
556 magnetic resonance imaging and multiple-breath washout. *Eur Respir J*  
557 2018;52:1800821. <https://doi.org/10.1183/13993003.00821-2018>.
- 558 [44] Smith L, Hughes PJ, Marshall H, Collier G, West N, Wildman M, et al.  
559 Hyperpolarised Gas MRI Shows Reversibility of Lung Ventilation Defects at  
560 Increased Inspiratory Lung Volumes in Cystic Fibrosis. A72. WHAT'S NEW  
561 Cyst. FIBROSIS, BPD OTHER Congenit. *Pediatr. LUNG Dis.*, San Diego:  
562 American Thoracic Society; 2018, p. A2313–A2313.
- 563 [45] Smith L, Marshall H, Aldag I, Horn F, Collier G, Hughes D, et al. Longitudinal  
564 Assessment of Children with Mild Cystic Fibrosis Using Hyperpolarized Gas  
565 Lung Magnetic Resonance Imaging and Lung Clearance Index. *Am J Respir*  
566 *Crit Care Med* 2018;197:397–400. [https://doi.org/10.1164/rccm.201705-](https://doi.org/10.1164/rccm.201705-0894LE)  
567 0894LE.
- 568 [46] Smith LJ, Marshall H, Bray J, Wildman M, West N, Horsley A, et al. The effect  
569 of acute maximal exercise on the regional distribution of ventilation using  
570 ventilation MRI in CF. *J Cyst Fibros* 2020.  
571 <https://doi.org/10.1016/j.jcf.2020.08.009>.
- 572 [47] van Kampen DA, Willems W, van Beers LWAH, Castelein RM, Scholtes VAB,  
573 Terwee CB. Determination and comparison of the smallest detectable change  
574 (SDC) and the minimal important change (MIC) of four-shoulder patient-  
575 reported outcome measures (PROMs). *J Orthop Surg Res* 2013;8:40.  
576 <https://doi.org/10.1186/1749-799X-8-40>.
- 577 [48] McNulty W, Usmani OS. Techniques of assessing small airways dysfunction.

- 578 Eur Clin Respir J 2014;1:25898. <https://doi.org/10.3402/ecrj.v1.25898>.
- 579 [49] Francisco B, Ner Z, Ge B, Hewett J, König P. Sensitivity of different spirometric  
580 tests for detecting airway obstruction in childhood asthma. *J Asthma*  
581 2015;52:505–11. <https://doi.org/10.3109/02770903.2014.984842>.
- 582 [50] Hughes PJC, Smith L, Chan H-F, Tahir BA, Norquay G, Collier GJ, et al.  
583 Assessment of the influence of lung inflation state on the quantitative  
584 parameters derived from hyperpolarized gas lung ventilation MRI in healthy  
585 volunteers. *J Appl Physiol* 2019;126:183–92.  
586 <https://doi.org/10.1152/jappphysiol.00464.2018>.
- 587 [51] He M, Driehuys B, Que LG, Huang Y-CT. Using Hyperpolarized <sup>129</sup>Xe MRI to  
588 Quantify the Pulmonary Ventilation Distribution. *Acad Radiol* 2016;23:1521–  
589 31. <https://doi.org/10.1016/j.acra.2016.07.014>.
- 590 [52] Oren O, Kebebew E, Ioannidis JPA. Curbing Unnecessary and Wasted  
591 Diagnostic Imaging. *JAMA* 2019;321:245.  
592 <https://doi.org/10.1001/jama.2018.20295>.
- 593
- 594
- 595
- 596
- 597

## Appendix

### Appendix 1

**Table 3:** Excluded Studies after Full-text Assessment and Exclusion Reasons

| Authors (Date)        | No. of multiple publications | Did not have sufficient information about the comparison between CF and healthy individuals. | Did not meet inclusion criteria | Did not assess hyperpolarised gas MRI reliability and validity or no information how it was assessed | Only included one CF patient and there was no sufficient information about the comparison between CF and healthy individuals |
|-----------------------|------------------------------|--|---------------------------------|--|--|
| Altes et al (2015)    | 1                            | *  |                                 | *  |  |
| Carlson et al (2018)  | None                         | *  |                                 | *  |  |
| Donnelly et al (1999) | None                         |  | *                               | *  |  |
| Horn et al (2014)     | 1                            |  | *                               | *  | *  |
| Kirby et al (2012)    | None                         |  | *                               | *  |  |
| Marshall et al (2014) | None                         | *  |                                 | *  |  |

|                         |      |   |   |   |
|-------------------------|------|---|---|---|
| Qing et al (2015)       | None |   | * | * |
| Smith et al (2017)      | None |   | * | * |
| Sun et al (2011)        | None | * |   | * |
| Thomen et al (2017)     | 1    | * |   | * |
| Tustison et al (2011)   | None |   | * | * |
| Youn et al (2012)       | None | * |   | * |
| Niedbalski et al (2019) | None |   | * | * |
| Munidasa et al (2019)   | None | * |   | * |

## Appendix 2

**Table 9:** Hypotheses Testing Findings for construct (convergent) validity

| Authors (Date) | Study hypothesis | Results | Results in support of hypothesis? |
|----------------|------------------|---------|-----------------------------------|
|----------------|------------------|---------|-----------------------------------|

|                           |   |  |     |
|---------------------------|---|--|-----|
| McMahon et al (2006) [4]  | " <sup>3</sup> He MRI would correlate with the major structural abnormalities seen on HRCT and also with functional information provided by spirometry, thus indicating a potential role as a marker of disease status in CF" | <sup>3</sup> He MRI had strong functional correlation with spirometry and structural CT abnormalities  | Yes |
| Bannier et al (2010) [22] | NA  | CF patients had ventilation defects, even though spirometry results showed normal lung function.   | NA  |
| Kirby et al (2011) [18]   | "He MRI would provide the necessary and sufficient spatial and temporal sensitivity to detect day-to-day changes in lung function"  | The results showed changes in ventilation defects when the <sup>3</sup> He-MRI repeated after 7 days, but day to day changes in the lung was not assessed. | No  |
| Altes et al (2012) [40]   | " <sup>3</sup> He-MRI would be appropriate for evaluating response to ivacaftor"  | <sup>3</sup> He MRI was able to detect the lungs response to ivacaftor which was effective in improving lung ventilation in CF patients                    | Yes |
| Smith et al (2018) [43]   | NA  | VDP strongly correlated with lung clearance index and forced expiratory volume in 1 s (FEV1)   | NA  |
| Zha et al (2019) [8]      | "Oxygen enhanced MRI may yield comparable whole-lung VDP relative to hyperpolarized <sup>3</sup> He MRI as the reference method"  | OE MRI showed similar performance compared with <sup>3</sup> He MRI for measuring VDP  | Yes |

|                          |    |  |    |
|--------------------------|----|--|----|
| Smith et al (2018) [45]  | NA | Ventilation MRI is capable of detecting significant lung function changes in the follow-up of children with CF and normal spirometry   | NA |
| Smith et al, (2019) [26] | NA | There was no inherent bias for VDP between the two gases although at an individual level differences were evident. Despite this, when followed up both gases similarly reflected changes in ventilation, suggesting both are capable of reflecting CF lung disease severity. | NA |

---

