

BMJ Open

BMJ Open is committed to open peer review. As part of this commitment we make the peer review history of every article we publish publicly available.

When an article is published we post the peer reviewers' comments and the authors' responses online. We also post the versions of the paper that were used during peer review. These are the versions that the peer review comments apply to.

The versions of the paper that follow are the versions that were submitted during the peer review process. They are not the versions of record or the final published versions. They should not be cited or distributed as the published version of this manuscript.

BMJ Open is an open access journal and the full, final, typeset and author-corrected version of record of the manuscript is available on our site with no access controls, subscription charges or pay-per-view fees (<http://bmjopen.bmj.com>).

If you have any questions on BMJ Open's open peer review process please email info.bmjopen@bmj.com

Assessment of Breath Volatile Organic Compounds in Acute Cardio-respiratory Breathlessness: A Prospective Real World Observational Study

Journal:	<i>BMJ Open</i>
Manuscript ID	bmjopen-2018-025486
Article Type:	Protocol
Date Submitted by the Author:	17-Jul-2018
Complete List of Authors:	<p>Ibrahim, Wadah; Leicester NIHR Biomedical Research Centre (Respiratory theme); College of Life Sciences, Department of Infection, Immunity and Inflammation, University of Leicester</p> <p>Wilde, Michael; University of Leicester, Department of Chemistry</p> <p>Cordell, Rebecca; University of Leicester, Department of Chemistry</p> <p>Salman, Dahlia; Loughborough University, Department of Chemistry</p> <p>Ruszkiewicz, Dorota; Loughborough University, Department of Chemistry</p> <p>Bryant, Luke; University of Leicester, Department of Chemistry</p> <p>Richardson, Matthew; Leicester NIHR Biomedical Research Centre (Respiratory theme); University of Leicester, Infection, Inflammation and Immunity</p> <p>Free, Robert; Leicester NIHR Biomedical Research Centre (Respiratory theme); University of Leicester, Infection, Inflammation and Immunity</p> <p>Zhao, Bo; Leicester NIHR Biomedical Research Centre (Respiratory theme); University of Leicester, Infection, Inflammation and Immunity</p> <p>Yousuf, Ahmed; University of Leicester, Infection, Inflammation and Immunity; Leicester NIHR Biomedical Research Centre (Respiratory theme)</p> <p>White, Christobelle; University of Leicester, Infection, Inflammation and Immunity; Leicester NIHR Biomedical Research Centre (Respiratory theme)</p> <p>Russell, Richard; University of Leicester, Infection, Inflammation and Immunity; Leicester NIHR Biomedical Research Centre (Respiratory theme)</p> <p>Jones, Sheila; University Hospitals of Leicester NHS Trust</p> <p>Patel, Bharti; University Hospitals of Leicester NHS Trust</p> <p>Awal, Asia; University Hospitals of Leicester NHS Trust</p> <p>Phillips, Rachael; NIHR Leicester Clinical Research Facility</p> <p>Fowkes, Graham; NIHR Leicester Clinical Research Facility</p> <p>McNally, Teresa; Paediatric Clinical Investigation Centre</p> <p>Foxon, Clare; Paediatric Clinical Investigation Centre</p> <p>Bhatt, Hetan; Leicester NIHR Biomedical Research Centre (Respiratory theme)</p> <p>Peltrini, Rosa; Leicester NIHR Biomedical Research Centre (Respiratory theme); University of Leicester, Department Infection, Inflammation and Immunity</p> <p>Singapuri, Amisha; Leicester NIHR Biomedical Research Centre (Respiratory theme)</p> <p>Hargadon, Beverley; Leicester NIHR Biomedical Research Centre (Respiratory theme); University of Leicester, Infection, Inflammation and Immunity</p> <p>Suzuki, T; Department of Cardiovascular Sciences, University of Leicester, , Cardiovascular Research Centre; Leicester NIHR Biomedical Research</p>

	<p>Centre (Cardiovascular theme) Ng, Leong; Department of Cardiovascular Sciences, University of Leicester, Cardiovascular Research Centre; Leicester NIHR Biomedical Research Centre (Cardiovascular theme) Gaillard, Erol; Leicester NIHR Biomedical Research Centre (Respiratory theme); Paediatric Clinical Investigation Centre Beardsmore, Caroline; Leicester NIHR Biomedical Research Centre (Respiratory theme); Paediatric Clinical Investigation Centre Ryanna, Kimuli; University Hospitals of Leicester NHS Trust Pandya, Hitesh; Discovery Medicine, Respiratory Therapeutic Area, GlaxoSmithKline PLC Coates, Tim; University of Leicester, Infection, Inflammation and Immunity; Department of Cardiovascular Sciences, University of Leicester Monks, Paul; University of Leicester, Infection, Inflammation and Immunity; University of Leicester, Department of Chemistry Greening, Neil; Leicester NIHR Biomedical Research Centre (Respiratory theme); University of Leicester, Infection, Inflammation and Immunity Brightling, Christopher; Leicester NIHR Biomedical Research Centre (Respiratory theme); University of Leicester, Infection, Inflammation and Immunity Thomas, Paul; Loughborough University, Department of Chemistry Siddiqui, Salman; Leicester NIHR Biomedical Research Centre (Respiratory theme); University of Leicester, Infection, Inflammation and Immunity</p>
Keywords:	Breathlessness, Observational study, Volatile Organic Compound, Breath analysis

SCHOLARONE™
Manuscripts

Review only

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

Assessment of Breath Volatile Organic Compounds in Acute Cardio-respiratory Breathlessness: A Prospective Real World Observational Study

Wadah Ibrahim^{1,2}, Michael Wilde³, Rebecca Cordell³, Dahlia Salman⁴, Dorota Ruszkiewicz⁴, Luke Bryant³, Matthew Richardson^{1,2}, Robert C. Free^{1,2}, Bo Zhao^{1,2}, Ahmed Yousuf^{1,2}, Christobelle White^{1,2}, Richard J Russell^{1,2}, Sheila Jones⁵, Bharti Patel⁵, Asia Awal⁵, Rachael Phillips⁶, Graham Fowkes⁶, Teresa McNally⁷, Clare Foxon⁷, Hetan Bhatt^{1,2}, Rosa Peltrini^{1,2}, Amisha Singapuri^{1,2}, Beverley Hargadon^{1,2}, Toru Suzuki^{8,9}, Leong L. Ng^{8,9}, Erol A Gaillard^{2,7}, Caroline Beardsmore^{2,7}, Kimuli Ryanna⁵, Hitesh Pandya¹¹, Tim Coates^{2,10}, Paul Monks^{2,3}, Neil Greening^{1,2}, Chris E Brightling^{1,2}, Paul Thomas⁴, Salman Siddiqui^{1,2}

wadah.ibrahim@leicester.ac.uk; mjw77@leicester.ac.uk; rc145@leicester.ac.uk; d.salman@lboro.ac.uk; d.m.ruszkiewicz2@lboro.ac.uk; lb368@leicester.ac.uk; mr251@leicester.ac.uk; rcf8@leicester.ac.uk; bo.zhao@leicester.ac.uk; ajy7@leicester.ac.uk; cmjw3@leicester.ac.uk; rjr22@leicester.ac.uk; rjr22@leicester.ac.uk; sheila.jones@uhl-tr.nhs.uk; bharti.m.patel@uhl-tr.nhs.uk; asia.awal@uhl-tr.nhs.uk; rachael.phillips@uhl-tr.nhs.uk; graham.fowkes@uhl-tr.nhs.uk; tm23@leicester.ac.uk; clare.foxon@uhl-tr.nhs.uk; hetan.bhatt@uhl-tr.nhs.uk; rp417@leicester.ac.uk; amisha.singapuri@uhl-tr.nhs.uk; beverley.hargadon@uhl-tr.nhs.uk; ts263@le.ac.uk; lln1@le.ac.uk; eag15@leicester.ac.uk; csb@leicester.ac.uk; kimuli.ryanna@uhl-tr.nhs.uk; hitesh.x.pandya@gsk.com; tim.coates@leicester.ac.uk; paul.monks@leicester.ac.uk; neil.greening@leicester.ac.uk; ceb17@leicester.ac.uk; C.L.P.Thomas@lboro.ac.uk; ss338@leicester.ac.uk

Corresponding author: Professor Salman Siddiqui, ss338@le.ac.uk

NIHR Leicester Biomedical Research Centre (Respiratory Theme)

Glenfield Hospital,

Groby Road

Leicester

LE3 9QP

t: +44 116 256 3841

e: ss338@le.ac.uk

Authors' affiliations:

¹Leicester NIHR Biomedical Research Centre (Respiratory theme), Glenfield Hospital, Groby Road, Leicester LE3 9QP

² College of Life Sciences, Department of Infection, Immunity and Inflammation, University of Leicester, University Road, Leicester LE1 7RH

³ Department of Chemistry, University of Leicester, University Road, Leicester, LE1 7RH

- 1
2
3 39 ⁴Department of Chemistry, Loughborough University, Epinal Way, Loughborough, LE11 3TU
- 4
5 40 ⁵ University Hospitals of Leicester NHS Trust, Glenfield Hospital, Groby Road, Leicester LE3 9QP
- 6
7 41 ⁶ NIHR Leicester Clinical Research Facility, Research Space, Level 0, Balmoral Building, Leicester Royal
8 42 Infirmery, Infirmery Square, Leicester, LE1 5WW
- 9
10 43 ⁷ Paediatric Clinical Investigation Centre, Leicester Respiratory Biomedical Research Centre, University of
11 44 Leicester, Leicester Royal Infirmery, Leicester, LE2 7LX
- 12
13 45 ⁸ Department of Cardiovascular Sciences, University of Leicester, Cardiovascular Research Centre, Glenfield
14 46 General Hospital, Leicester LE3 9QP.
- 15
16 47 ⁹ Leicester NIHR Biomedical Research Centre (Cardiovascular theme), Glenfield Hospital, Groby Road,
17 48 Leicester LE3 9QP
- 19
20 49 ¹⁰Department of Cardiovascular Sciences, University of Leicester, Level G, Jarvis building RMO, Infirmery
21 50 Square, Leicester LE1 5WW
- 22
23
24 51 ¹¹ Discovery Medicine, Respiratory Therapeutic Area, GlaxoSmithKline PLC, Gunnels Wood Road, Stevenage,
25 52 Hertfordshire, SG1 2NY
- 26
27 53
28 54
29 55
30 56
31 57
32 58
33 59
34 60
35 61
36 62
37 63
38 64
39 65
40 66
41 67
42 68

69 **Abstract**

70 **Introduction**

71 Patients presenting with acute undifferentiated breathlessness are commonly encountered in
72 admissions units across the United Kingdom. Existing blood biomarkers have clinical utility in
73 distinguishing patients with single organ pathologies but have poor discriminatory power in multi
74 factorial presentations. Evaluation of volatile organic compounds (VOC) in exhaled breath offers the
75 potential to develop biomarkers of disease states that underpin acute cardio-respiratory breathlessness,
76 owing to their proximity to the cardio-respiratory system. To date there has been no systematic
77 evaluation of VOC in acute cardio-respiratory breathlessness. The proposed study will seek to use
78 both offline and online VOC technologies to evaluate the predictive value of VOC in identifying
79 common conditions that present with acute cardio-respiratory breathlessness.

80 **Methods and analysis**

81 A prospective real world observational study carried out across three acute admissions units within
82 Leicestershire. Participants with self-reported acute breathlessness, with a confirmed primary
83 diagnosis of either acute heart failure, community acquired pneumonia and acute exacerbation of
84 asthma or COPD will be recruited within 24 hours of admission. Additionally, school age children
85 admitted with severe asthma will be evaluated. All participants will undergo breath sampling on
86 admission and upon recovery following discharge. A range of online technologies including: proton-
87 transfer-reaction mass spectrometry (PTR-MS), gas chromatography ion mobility spectrometry (GC-
88 IMS), atmospheric pressure chemical ionisation- mass spectrometry (APCI-MS) and offline
89 technologies including gas chromatography mass spectroscopy (GC-MS) and comprehensive two-
90 dimensional gas chromatography-mass spectrometry (GCxGC-MS) will be utilised for VOC
91 discovery and replication. For offline technologies a standardised CE marked breath sampling device
92 (ReCIVA[®]) will be used. All recruited participants will be characterised using existing blood
93 biomarkers including C - reactive protein (CRP), brain derived natriuretic peptide (BNP), Troponin-I
94 and blood eosinophil levels and further evaluated using a range of standardised questionnaires, lung
95 function testing, sputum cell counts and other diagnostic tests pertinent to acute disease.

1
2
3 96 **Ethics and dissemination**

4
5 97 The National Research Ethics Service Committee East Midlands has approved the study protocol
6
7 98 (REC number: 16/LO/1747). IRAS 198921. Findings will be presented at academic conferences and
8
9 99 published in peer-reviewed scientific journals. Dissemination will be facilitated via a partnership with
10
11 100 the East Midlands Academic Health Sciences Network and via interaction with all UK funded
12
13 101 MRC/EPSRC molecular pathology nodes.
14
15
16 102

17
18 103 **Key words:** Breathlessness; Breath analysis; Volatile Organic Compound, Observational study
19
20 104
21
22 105
23
24 106
25
26 107
27
28 108
29
30 109
31
32 110
33
34 111
35
36 112
37
38 113
39
40 114
41
42 115
43
44 116
45
46 117
47
48
49
50
51
52
53
54
55
56
57
58
59
60

118 **Strengths and Limitations of this Study**

119 **Strengths**

- 120 • A pragmatic real world, prospective, observational study across three admission units that
121 focuses on the systematic discovery and replication of VOC in acutely breathless patients
122 using both online and offline technologies
- 123 • The study will evaluate populations that often present with diagnostic uncertainty
124 including elderly multi-morbid patients and school age children
- 125 • The proposed study is the largest of its kind in acute disease to characterise VOC with a
126 range of additional assessments that will build a comprehensive phenotype of acute
127 cardio-respiratory exacerbations
- 128 • Benchmarking of candidate VOC to established blood based biomarkers utilised in
129 clinical practice e.g. BNP, CRP will form an important method of validation against
130 existing molecular pathology
- 131 • The proposed study will build an infrastructure for research and subsequent evaluation of
132 VOC in interventional trials within acute cardio-respiratory exacerbations

133 **Limitations**

- 134 • Prior acute treatment exposure will need to be accounted for when evaluating potential
135 discriminative biomarkers
- 136 • VOC technologies are not currently suited for deployment in patients that are of high clinical
137 acuity e.g. severe respiratory failure and thus the study will self-select patients that can be
138 safely sampled with online and offline methodologies
- 139 • Although the study will quantify and evaluate diagnostic uncertainty, the study does not
140 enrich for these cases and the patient population will predominantly comprise of patients in
141 whom a senior clinical decision maker could make the primary clinical diagnosis within 24
142 hours of admission. As a consequence future studies will be required to evaluate the
143 biomarkers developed in cardio-respiratory exacerbations where the diagnosis is unclear.

144 **1. Introduction:**

145 Breathlessness is a common symptom of cardio-respiratory illnesses that has a significant direct
146 impact on patients' wellbeing as well as a substantial economic burden on healthcare systems [1].

147 Although its etiologies can be variable, exacerbations of common complex chronic cardio-respiratory
148 conditions account for approximately 70% of acute presentations with breathlessness, namely
149 exacerbations of asthma and COPD, acute heart failure and community acquired pneumonia [2].

150 Moreover, moderate and severe breathlessness is significantly associated with all-cause,
151 cardiovascular and COPD mortality[3]. As a consequence symptomatic breathlessness warrants rapid
152 evaluation and targeted diagnostics at presentation.

153

154 Diagnostic evaluation of acute breathlessness is heavily reliant on blood based biomarkers e.g. CRP,
155 BNP, Troponin and on occasions blood eosinophil levels. These biomarkers have clinical utility
156 primarily in patients with single pathologies, but have poor discriminatory power in patients with
157 multifactorial presentations of acute breathlessness[4]. There is therefore an unmet need for the
158 development of sensitive and specific biomarkers that differentiate acute breathlessness from its
159 recovery and the common cardio-respiratory conditions that present with acute breathlessness.

160

161 CRP plays an important role in diagnosing breathlessness caused by an underlying bacterial
162 pneumonia[5], as well as predicting mortality in patients with COPD [6]. BNP is routinely utilised in
163 acute settings to support the diagnosis of acute heart failure [7].The European Society of Cardiology
164 (ESC) recommends BNP threshold values of <100 pg/mL to rule out acute congestive cardiac failure
165 and values > 500 pg/ml as diagnostic of acute exacerbations of heart failure [8].

166

167 The role of peripheral blood eosinophil count in airway inflammation was poorly understood up until
168 the second half of the 19th century when Paul Ehrlich, a German physician and Nobel prize winner,
169 introduced eosin in his technique for white cell differentiation in 1879[9] . Considerable advances in
170 the field of airway inflammation and the role of eosinophils have taken place since [10-12]. More

1
2
3 171 recently Bafadhel *et al* suggested that peripheral blood eosinophil count can be used to direct
4
5 172 corticosteroid therapy during COPD exacerbations in single centre study [13].
6
7 173 Currently, blood biomarkers together with clinical, physiological and imaging parameters are used in
8
9 174 diagnosing the cause of acute breathlessness. Blood biomarkers may be less specific as they originate
10
11 175 far from the target organs of interest (the heart and the lungs in cardio-respiratory disease). Sputum,
12
13 176 although potentially a more definitive lung specific matrix, is comparatively difficult to obtain
14
15 177 particularly in acutely unwell patients, limiting its use in acute disease and highlighting the need for
16
17 178 better biomarkers. Ideally these biomarkers would have the following characteristics, (i) they would
18
19 179 originate from the target organ of interest, (ii) they would significantly add value to conventional risk
20
21 180 scoring and diagnostic algorithms in acute breathlessness, (iii) they would be minimally invasive and
22
23 181 suitable for rapid point of care diagnosis in emergency rooms and acute admissions units (iv) they
24
25 182 would have diagnostic value in patients with multifactorial acute breathlessness.

26
27 183
28 184 Exhaled breath contains thousands of volatile organic compounds (VOC) that reflect biological
29
30 185 processes occurring in the host both locally in the airways and systematically offering the potential to
31
32 186 develop more effective biomarkers in acutely breathless patients (**Figure 1**).

33
34 187
35
36 188 The proposed program of research will use a combination of offline and online technologies to
37
38 189 identify and evaluate the diagnostic and prognostic value of VOC in patients with acute cardio-
39
40 190 respiratory related breathlessness (**Figure 2**).

41
42 191
43
44 192 Exhaled breath analysis of volatile chemicals offers the potential to develop more effective biomarkers
45
46 193 in acutely breathless patients. The use of breath analysis for disease diagnostics dates back to ancient
47
48 194 Greeks where physicians used exhaled breath to diagnose different diseases. Breath odours allow
49
50 195 correct associations to certain diseases. For example, the sweet smell of diabetic ketoacidosis, the
51
52 196 fishy smell of breath associated to liver illness, the urine-like odour of kidney disease and the smell of
53
54 197 the breath of patients with lung abscesses, caused by the proliferation of anaerobic bacteria [14-17].

1
2
3 198 More recently, exhaled breath analysis has demonstrated early proof of concept in the diagnosis of
4
5 199 acute heart failure, ventilator associated pneumonia[18] and stable state airways disease [15]. The
6
7 200 validity of breath analysis has also been demonstrated in breathless children[19]. This population is
8
9 201 likely to prefer breath-based tests, as these are minimally invasive. Importantly, a variety of point of
10
11 202 care sensors are now available to evaluate potential exhaled breath biomarkers in emergency care
12
13 203 settings.

14 204

15
16 205 Despite the novelty of non-invasive sampling technology and the growing interest in exhaled breath
17
18 206 analysis there remains a disappointing level of comparability across studies due to the lack of
19
20 207 standardisation and appropriate data analysis methods.

21
22 208 A recent systemic review by Anders Christiansen *et al* compared eleven publications reporting very
23
24 209 heterogeneous designs, methods, patient group sizes, data analytics and, consequently, quite varying
25
26 210 results [20].

27 211

28
29
30 212 To our knowledge, no other large studies exploring the use of breath biomarkers in profiling acute
31
32 213 breathlessness have been completed. Few studies have explored the use of electronic nose in stable
33
34 214 disease with good discriminatory power in COPD [21], Pneumonia [22] and heart failure[23] with
35
36 215 relatively small sample size. The focus of the current research study will be to evaluate acutely
37
38 216 breathless cardio-respiratory patients using a combination of ‘discovery’ and near-patient care breath
39
40 217 sampling technologies.

41
42 218 Medical Research Council (MRC) and Engineering and Physical Sciences Research Council (EPSRC)
43
44 219 have commissioned a series of molecular pathology nodes aimed at developing molecular signatures
45
46 220 relevant to disease diagnosis and progression. This was triggered by the clear need for alliance
47
48 221 between academic institutions, industry and NHS partners to enhance the benefits of stratified
49
50 222 medicine for patients[24, 25].

51
52 223 University of Leicester and Loughborough University were awarded a joint molecular pathology node
53
54 224 East Midlands Breathomics Pathology Node (EMBER) which this study forms a key part of.

55 225

226 **2. Methods and Analysis**

227 **2.1. Study design**

228 A prospective real world observational study across three acute admissions units within Leicestershire
229 (two adult admissions units and one children's assessment unit). The acute units routinely assess and
230 treat cardio-respiratory admissions due to breathlessness in adults and children.

231 Participants with self-reported acute breathlessness, either requiring admission or a change in baseline
232 treatment, will be screened for the study. Informed consent will be obtained in all participants
233 following a clinical review by a senior decision maker within 24 hours of acute admission (**Figure 3**).

234

235

236 **2.2. Objectives**

237 **2.2.1. Primary objective**

- 238
- 239 • To evaluate the sensitivity, specificity, positive and negative predictive value of exhaled
240 breath VOC biomarkers to differentiate acute breathlessness in cardio-respiratory patients.

241 **2.2.2. Secondary objectives:**

- 242
- 243 • To replicate selected breath VOC biomarkers identified in acute breathlessness.
- 244 • To discover and replicate breath VOC biomarkers that differentiate the common cardio-
245 respiratory conditions that cause acute breathlessness, specifically (i) acute heart failure, (ii)
246 community acquired pneumonia, (iii) adult exacerbations of asthma and COPD and age-
247 matched adults that do not have cardio-respiratory disease or breathlessness.
- 248 • To quantify the level of clinical uncertainty in the primary diagnosis using a 100 mm visual
249 analogue scale and independent clinical adjudication of case notes blinded to the following
250 blood biomarkers (i) CRP (ii) BNP (iii) Troponin-I (iv) blood eosinophils, but not clinical
251 history and acute presentation nor chest x-ray imaging. Potential discriminatory breath VOC
252 biomarkers will be adjusted for clinical uncertainty in statistical models.
- 253 • To identify and replicate exhaled breath VOC biomarkers in school age children treated in
254 hospital for severe asthma attacks and compare these to age-matched healthy controls.

1
2
3 255 **2.2.3. Exploratory end points (where applicable):**
4 256

- 5 257 • To evaluate the dynamic profile of selected breath VOC between the acute state and the
6
7 258 recovery state post exacerbation.
8
9 259 • To evaluate the relationship between exhaled VOC biomarkers and clinical outcomes
10
11 260 including (i) hospital readmission at 30 and 60 days post event (ii) all-cause mortality over a 2
12
13 261 year period post admission.
14
15 262 • To evaluate the relationship between breath VOC biomarkers and functional measures
16
17 263 e.g. physical performance and activity
18
19 264 • To explore potential breath VOC biomarkers of multifactorial acute breathlessness
20
21 265 • To evaluate the relationship between diet, lifestyle and environment upon breath VOC
22
23 266 biomarkers
24

25
26 267
27 268 **2.3. Sample size estimation**
28

29 269
30
31 270 Preliminary data was utilised to conduct sample size estimates from a cohort of acutely breathless
32
33 271 patients admitted to acute admissions units over a 6 month period (February 2017 to August 2017).
34
35 272 Hundred and twelve adult participants (asthma 46, community acquired pneumonia 26, COPD 22)
36
37 273 and eighteen healthy controls were utilised for the analysis.

38
39 274 A panel of ten pre-specified aldehydes, based on literature search [26-28], were extracted from breath
40
41 275 using gas chromatography-mass spectrometry (GC-MS). The aldehydes were normalised to a
42
43 276 common internal standard and were not background-subtracted.
44

45
46 277 A closed formula from Hsieh *et al*[29], relating sample size to observable effect size, was used to
47
48 278 calculate sample size from logistic regression models of the ten aldehydes with acute breathlessness
49
50 279 as the outcome measure. The sample size estimates are also relevant to acute class comparisons vs.
51
52 280 the sum of other acute classes.

53
54 281 Based upon the sample size estimates we would have an 80% power, with a type 1 error rate of 5%, to
55
56 282 detect an odds ratio of association of 1.2 between two disease classes with 55 patients per class. Given
57

283 the fact that study seeks to discover and replicate breath VOC amongst five adult disease classes
284 (community acquired pneumonia, heart failure, COPD, asthma and healthy aged matched subjects) we
285 would require 110 adult patients per class – 550 patients across the program to achieve these aims.

286 The closed formulae by Tihaki *et al*,[30] were also utilised to understand the discriminatory power
287 that the samples sizes above would provide with respect to biomarker sensitivity and specificity; The
288 following assumptions were made:

- 289 • That a sensitivity of 80 % with a precision of 5% would provide a useful biomarker capable
290 of ‘ruling out’ an acute class. The same target was applied to specificity.
- 291 • We assume a prevalence of acute breathlessness of 80% as the recruitment campaign uses
292 acute breathlessness as the initial stratification tool for recruitment and 1:5 patients recruited
293 will be non-breathless healthy controls
- 294 • We aim to balance group sizes across classes equally

295 For a type 1 error rate of 0.05 and a 95% confidence interval

296 $N_{\text{sensitivity}} = 307$

297 $N_{\text{specificity}} = 1,230$

298

299 For a type 1 error rate of 0.05 and a 90% confidence interval

300 $N_{\text{sensitivity}} = 218$

301 $N_{\text{specificity}} = 871$

302

303 For a type 1 error rate of 0.05 and an 85% confidence interval

304 $N_{\text{sensitivity}} = 166$

305 $N_{\text{specificity}} = 664$

306

307 For a type 1 error rate of 0.05 and an 80% confidence interval

308 $N_{\text{sensitivity}} = 131$

309 $N_{\text{specificity}} = 524$

310

1
2
3 311 Therefore, we are powered to identify sensitive biomarkers ($\geq 80\%$) of acute breathlessness with a
4
5 312 maximum marginal error in the estimate for sensitivity not exceeding 5% with 95% confidence.
6
7 313 Similarly, we are powered to identify specific biomarkers ($\geq 80\%$) of acute breathlessness with a
8
9 314 maximum marginal error in the estimate for specificity not exceeding 5% with 80% confidence.
10
11 315 For the primary analysis the outcome will be treated as a nominal variable with levels (i) acute heart
12
13 316 failure (ii) community acquired pneumonia (iii) adult exacerbations of asthma and COPD and (iv)
14
15 317 acute exacerbations in school age children treated in hospital for severe asthma attacks.
16
17
18 318 The relationship between the primary outcome and the exhaled breath VOC biomarkers will be
19
20 319 modelled using multinomial logistic regression. In addition to metabolomics markers the following
21
22 320 independent variables will be included in the model: clinical uncertainty score on a 100 mm VAS
23
24 321 scale, age, and a validated co morbidity score (the Charlson comorbidity score)[31, 32].
25
26 322
27
28 323 Receiver operator analyses will be utilised to generate ROC curves for individual and multiple panels
29
30 324 VOC predictors in the primary analysis.
31
32 325
33
34 326 To understand the dynamic profile of breath biomarkers during (i) the acute state and (ii) in the
35
36 327 chronic state up to 6 months post exacerbation, a repeated measures model with a random intercept
37
38 328 and random effect for time will be fitted, the random effects will be fitted for each patient. For the
39
40 329 repeated measures mixed model an unstructured covariance will be assumed. To evaluate the
41
42 330 relationship between breath biomarkers and hospital readmission at 30 and 60 days Cox proportional
43
44 331 hazards and frailty models will be utilised [33]. Analysis of Multivariate Survival Data, [CITE]
45
46 332 competing risk models and joint models will be fitted [34]. Relationship between death and breath
47
48 333 biomarkers will be evaluated using a logistic regression model. Changes in outcome measures will be
49
50 334 measured appropriately for each variable (e.g. paired t-test, Mann-Whitney, repeated measures
51
52 335 analysis). Tables of descriptive statistics will be compiled for all key variables
53
54
55 336 All analysis will be performed using R 3.5.0 <https://www.r-project.org/>.
56
57 337

338 2.4. Discovery and Replication studies

339 Specific indicator conditions have been selected for targeted recruitment according to their high
340 prevalence and unmet need, their high morbidity and mortality and the need to develop better
341 diagnostic and prognostic algorithms in acute care pathways.

342 The indicator diagnoses of interest are (i) exacerbations of adult asthma and COPD, (ii) community
343 acquired pneumonia (iii) acute heart Failure and (iv) exacerbation in school age children treated in
344 hospital for severe asthma attacks.

345 Patient level clinico-pathological and outcome data (spanning the entire acute pathway) will be
346 collected in parallel to breath sampling. In addition, breath samples will be acquired in the stable state
347 post exacerbation (**Figure 3**).

348
349 Age matched healthy volunteers will be recruited where possible at separate visits. For acute admission
350 the study team will approach the spouse, parent or sibling of the index case and seek informed consent
351 for study assessments. All healthy subjects will undergo two assessments separated by a duration of 8-
352 16 weeks to match the acute and recovery time points elapsed in their index
353 case/partner/spouse/sibling/child. Additional healthy volunteers will be identified from local
354 recruitment databases and via advertising

356 2.4.1. Discovery Phase (Project months 1-24):

357 The aim of the discovery phase is to identify putative discriminatory breath VOC, using both offline
358 and online technologies.

359 Pre-planned recruitment of acutely breathless patients will be enriched into the following disease
360 strata following senior clinical decision maker assessment and within 24 hours of acute admission.

361 Acute adult heart failure (n=55), adult community acquired pneumonia (n=55), adult exacerbations of
362 asthma (n=55) and COPD (n=55), acute severe asthma attacks in school age children (n=55).

363 Additional age matched healthy volunteers (n=55 adults and 50 children) will be identified as a non-
364 disease reference group.

1
2
3 365 **2.4.2. Replication Phase (years 3-4)**

4 366 The aim of the replication phase is to replicate putative discriminatory breath VOC /VOC signatures
5
6 367 identified in the discovery phase.

7
8 368 The aim of the discovery phase is to identify putative discriminatory breath VOC, using both offline
9
10 369 and online technologies.

11
12 370 Pre-planned recruitment of acutely breathless patients will be enriched into the following disease
13
14 371 strata following senior clinical decision maker assessment and within 24 hours of acute admission.

15
16 372 Acute adult heart failure (n=55), adult community acquired pneumonia (n=55), adult exacerbations of
17
18 373 asthma (n=55) and COPD (n=55), acute severe asthma attacks in school age children (n=55).

19
20 374 Additional age matched healthy volunteers (n=55 adults and 50 children) will be identified as a non-
21
22 375 disease reference group.

23
24 376 **Total combined sample size of the discovery and replication phases = 650 participants**

25
26 377

27
28 378 **2.5. Schedule of assessments**

29
30 379 A schedule of acute assessments is outlined below and aligns to the movement of acute patients
31
32 380 through the clinical care pathway and the overall aim of developing a complete phenotypic picture of
33
34 381 acutely breathless patients.

35
36 382

37
38 383 **2.5.1. Defining acute breathlessness**

39
40 384 At presentation (within 24 hours of admission) to one of three acute admissions units potentially
41
42 385 eligible patients will be identified following confirmation of acute breathlessness, identified as (i)
43
44 386 patient defined acute breathlessness and/or (ii) 1 unit increase above patient reported baseline in the
45
46 387 extended medical research council (eMRC) dyspnoea score [35, 36] and at least one of the indicator
47
48 388 diagnoses identified as the primary clinical diagnosis by a senior clinical decision maker.

49
50 389

51
52 390

53
54 391

392 2.5.2 Informed consent

393 Patients meeting the pre-specified definition of acute breathlessness will be approached for informed
394 consent in to the breath VOC biomarker study. Only patients that are eligible to give full written
395 informed consent will be recruited.

397 2.5.3 Collection of blood based pathology markers

398 Collection of the blood biomarkers CRP, BNP, Troponin-I and blood eosinophil count will be
399 performed both acutely and following recovery, when not taken as part of clinical care pathway.

400 These are currently used in profiling acutely breathless patients in clinical practice (**Table1**).

Test	ANALYSER/METHOD	LOWER LIMIT OF DETECTION	UPPER LIMIT OF DETECTION
C-Reactive protein (CRP)	Siemens Advia Chemistry XPT, PEG enhanced immunoturbidimetric. Siemens Advia 1800, PEG enhanced immunoturbidimetric	5 mg/L	Diluted to result
B-type natriuretic peptide (BNP)	Siemens Advia Centaur XPT, two-site sandwich immunoassay using direct chemiluminescent technology	2.0 pg/mL	1445 pg/mL
Troponin-I	Abbott Architect i2000SR, three-site sandwich immunoassay using direct chemiluminescent technology (CMIA).	5.0 ng/L	50,000 ng/L

401

402 **Table (1):** Type of analyser and methodology used for blood biomarker calculation. The table
403 outlines analyser make, methodology, upper and lower limits of detection as per the University
404 Hospitals of Leicester NHS Foundation trust laboratory guidelines.

405

406 2.5.4 Breath VOC sampling

407 Offline breath sampling using GC-MS coupled with a standardised and CE marked breath sampler-
408 ReCIVA[®][37] and comprehensive two-dimensional gas chromatography-mass spectrometry, coupled
409 with a standardised and CE marked breath sampler (ReCIVA[®] GCxGC-MS) will be performed .
410 Additionally the following online technologies, proton transfer mass spectrometry (PTR-MS), gas
411 chromatography - ion mobility spectrometry (GC-IMS) and atmospheric pressure chemical ionisation-

412 mass spectrometry (APCI-MS) will be evaluated according to the sampling strategy outlined in

	COPD		Asthma		Pneumonia		Heart Failure		Healthy		Paediatrics	
Time point	1	2	1	2	1	2	1	2	1	2	1	2
Written informed consent	x		x		x		x		x		x	
Volatile organic compound (VOC) sampling												
ReCIVA- gas chromatography and mass spectrometry (GC-MS)	x	x	x	x	x	x	x	x	x	x	x	x
ReCIVA- comprehensive two-dimensional gas chromatography (GCxGC-MS)	x	x	x	x	x	x	x	x	x	x	x	x
Atmospheric pressure chemical ionisation- mass spectrometry (APCI-MS)	x	x	x	x	x	x	x	x	x	x	x	x
Proton transfer reaction mass spectrometry (PTRMS)		x		x		x		x	x	x	x	x
Gas chromatography - ion mobility spectrometry (GC-IMS)	x		x		x		x				x	x
Pathology blood tests												
Full blood count (including differential cell count)	x	x	x	x	x	x	x	x	x	x	x	x
Brain natriuretic peptide (BNP) [pg/mL]	x	x	x		x		x	x	x			
Troponin-I [ng/L]	x		x		x		x		x			
C-Reactive protein (CRP) [mg/L]	x	x	x	x	x	x	x	x	x	x	x	x
Lung function tests											x	x
Hand held forced oscillation technique (FOT)	x	x	x	x	x	x	x	x	x	x		
Fractional exhaled nitric oxide (FeNO) - Flow rate 50 [ml/s]			x	x							x	x
Spontaneous sputum sample	x	x	x	x	x	x	x	x	x	x		
Bio-banking (urine, serum, plasma. sputum supernatants & plugs)	x	x	x	x	x	x	x	x	x	x	x	x
Transthoracic echocardiography	x						x					

413 section 3, **Figure 3 and Table 2.**

414 **Table (2):** Summary of baseline and follow up assessments. The table summarises key assessments
 415 carried out at different time points during the study. The participants may undertake any combination
 416 of the investigations listed at any of these time points.

417

418

419

420

421 **2.5.5. Collection of additional samples for future biomarker campaigns**

422 Collection of additional biomarkers for future biomarker discovery campaigns including (i) a urine
423 sample, (ii) blood samples, up to 85mls for DNA, RNA, plasma and serum and peripheral blood cell
424 flow cytometry in selected subjects and (iii) spontaneous or induced sputum samples (plugs and
425 supernatants) will be carried out (**Table 2**).

426

427 All samples will be collected at time point 1 and at 2 (**Figure 3**). The additional samples will be used
428 for future omics analyses, these may include detailed analysis of the metagenome in sputum and
429 proteomics applied to urine and serum samples.

430

431 **2.5.6. Physiological characterisation**

432 Physiological measures of lung function will be performed in acutely ill participants and at
433 recovery including, (i) Hand held forced oscillation technique (FOT): an easily accessible
434 measure of lung function. Patients favour this to spirometry as it is effort independent, unlike
435 spirometry and requires less than a minute of quiet tidal breathing to obtain triplicate high
436 quality measurements [38], (ii) Fractional exhaled nitric oxide (FeNO): A measure of airway
437 inflammation in asthmatic patients[39, 40] (iii) Echocardiography: Two dimensional
438 transthoracic echocardiography was performed in heart failure and COPD patients using an
439 iE 33 system with S5-1 transducer (frequency transmitted 1.7 MHz, received 3.4 MHz;
440 Philips Medical Systems, Best, The Netherlands). Standard techniques as per American
441 Society of Echocardiography guidelines (ASE)[41] were used to acquire 2D, colour and
442 Doppler images in conventional parasternal long-axis, short-axis and apical 4-, 2- and 3-

1
2
3 443 chamber views. Left ventricular ejection fraction (LVEF) was calculated using the biplane
4
5 444 method of discs formula (Simpson's rule) to derive left ventricular volume indices.
6
7

8 445

10 446 **2.5.7 Recovery follow up**

11 447 The recovery from an acute exacerbation will be confirmed and identified as patient defined recovery,
12
13 448 at the recovery study visit (time point 2) up to six months post-acute event. The schedule of
14
15 449 assessments at the recovery visit is outlined (**Table 2**).
16
17

18 450

19 451 **2.6. Clinical Adjudication:**

20 452

21 453 In an effort to reduce data variability and minimise bias, an independent panel consisting of two senior
22
23 454 acute clinicians (SS & NG) will review all pertinent clinical and diagnostic source documentation,
24
25 455 whilst blinded to admission blood biomarkers and clinical diagnosis.
26

27 456 All acutely breathless adult patients' notes will be sequentially divided for adjudication. The panel will
28
29 457 independently determine the primary diagnosis of highest probability from a list of the four potential
30
31 458 acute indicator diagnoses in section 2.4 and mark their level of clinical certainty on a 100 mm visual
32
33 459 analogue scale (VAS scale). The panel members will be able to review imaging, electrocardiograms
34
35 460 (ECGs), and other relevant information but not admission blood based pathology tests.
36

37 461 In a subset of patients adjudication will be validated by separate panel member to ensure between
38
39 462 observer agreement using Bland-Altman analysis and inter-rater agreement of the primary diagnosis
40
41 463 using Kohen's kappa via repeated evaluation of a subset of cases (see statistical methods section 2.3).
42

43 464

45 465 **2.7. Clinical Informatics:**

46 466

47
48 467 Clinical data collection will be undertaken using a securely-hosted bespoke database system (ADD)
49
50 468 developed within the NIHR Leicester Biomedical Research Centre – Respiratory (BRC). The system
51
52 469 links acute admission episodes to hospital pathology records; historical respiratory physiology tests;
53
54 470 and demographic information. The system provides functionality to validate data entry; manually
55
56 471 verify records and highlight incomplete records. A custom VOC 'module' has been created to
57

1
2
3 472 support data collection within the study visits (1 and 2), and standardise diagnoses and medications
4
5 473 through the use of clinical ontologies as well as linking hospital records/tests to patient visits.
6
7 474
8
9 475 Non-ADD based clinical data (e.g. hospital admissions, re-admissions and mortality) will be extracted
10
11 476 from the hospital data warehouse using identifiable patient identifiers, and subsequently
12
13 477 pseudonymised prior to integration.
14
15 478

16
17 479 An informatics pipeline will be created to facilitate the transfer of chemo-metric data from remote
18
19 480 computers to the data repository. This will include tools to (i) enforce the correct labelling of data sets
20
21 481 (e.g. study number, visit, type/source of sample) prior to automated validated transfer to the
22
23 482 repository; (ii) record information about the sample process; (iii) search and extract data sets from the
24
25 483 repository for subsequent analysis. Prior to analysis, clinical and chemo-metric data will be integrated
26
27 484 using the study number and any potentially identifiable information will be removed.
28
29 485

30 486 **3. Breath profiling**

31 487
32
33 488 The technologies utilised in the VOC study during discovery and replication phases are:

34 35 36 489 Offline technologies

- 37 490 - ReCIVA+ GC-MS
- 38 491 - ReCIVA + GC x GC-MS
- 39 492 -

40 41 42 43 493 Online technologies

- 44 494 - GC-IMS
- 45 495 - PTR-MS
- 46 496 - APCI-MS

47
48
49
50 497
51
52 498 Offline technologies will underpin the discovery analyses owing to their ability to identify chemical
53
54 499 identity and their recognition as the analytical gold standard in exhaled breath VOC analysis [42].
55
56
57
58
59

1
2
3 500 In contrast online technologies will be utilised for VOC biomarker replication and at the recovery
4
5 501 visits owing to their portability and potential for future point of care testing. (Figure 4).

6
7 502 A brief description of the core VOC platforms is provided below

8
9 503 A CE marked breath sampling device (ReCIVA®) developed by Owlstone Medical, will be used to
10
11 504 sample breath onto adsorbent Tenax tubes. This effectively allows de-coupling of the breath sampling
12
13 505 from the breath sensor and analysis platforms in selected patients that are not able to mobilise to a real
14
15 506 time breath sampling device. The Owlstone ReCIVA sampler will be utilised in breath collection for
16
17 507 offline technologies namely GC-MS and GCxGC-MS. The ReCIVA sampler is capable of entraining
18
19 508 oxygen and is therefore suitable for patients with mild respiratory failure requiring low flow rates of
20
21 509 oxygen to maintain target oxygen saturations [37].

22
23
24 510 **3.1. Gas chromatography and mass spectroscopy (GC-MS):** is a commonly applied methodology
25
26 511 used to accurately measure trace gases in complex mixtures such as exhaled air [42]. Pre-
27
28 512 concentrating breath volatiles by various means and subsequent analysis constitute a reliable and
29
30 513 sensitive method for VOC analysis [43]. Despite its high sensitivity, it is however, a time consuming
31
32 514 technique and carries a risk of contamination at the pre-concentration step. It is also not suitable for
33
34 515 online and multiple measurements limiting its use as a point-of-care testing technology for VOC [44].

35
36 516 **3.2. Comprehensive two-dimensional gas chromatography-mass spectrometry (GCxGC-MS):** an
37
38 517 advanced analytical technique for the analysis of complex organic matrices; its main advantage is the
39
40 518 unparalleled separation power it affords over conventional one-dimensional chromatographic
41
42 519 techniques [45]. Previous research, albeit sparse, has demonstrated the potential of GCxGC-MS for
43
44 520 breath analysis with the number of VOC detected exceeding those detected by conventional GC-MS
45
46 521 [46, 47]. GCxGC-MS of breath metabolites has been used for the identification of biomarkers related
47
48 522 to glucose metabolism [48, 49], tuberculosis [50] and radiation response [51]. This has generated
49
50 523 interest within the breath research community, however, such studies were conducted on a small scale
51
52 524 (<50 patients) and involved the use of expensive detectors and modulators. Method development and
53
54 525 analysis of the data-rich GCxGC chromatograms, however, can be time-consuming and require
55
56 526 specialist knowledge.

1
2
3 527 **3.3. Proton-transfer-reaction mass spectrometry (PTRMS):** a real time technique, capable of
4
5 528 simultaneously measuring the evolution of multiple gas metabolites from a single breath. It has been
6
7 529 used for the identification of potential useful VOC biomarkers for diagnosis of a variety of diseases:
8
9 530 including various cancers[52-54], liver disease[55, 56] and respiratory disease[57]. It has several
10
11 531 advantages in clinical settings, such as the speed of sampling, the instant result achieved and the lack
12
13 532 of need for sample storage or shipping. However, owing to the lack of pre-concentration or
14
15 533 chromatographic separation, sensitivity and definitive compound identification can be somewhat
16
17 534 limited when compared to GC-MS.

18
19 535 **3.4. Gas chromatography- ion mobility spectrometry (GC-IMS):** allows the detection of volatile
20
21 536 organic compounds down to ultra-trace level (ng/L - to pg/L - range). For years, IMS has been used to
22
23 537 discover potential discriminatory breath VOC in lung cancer [58, 59], COPD[60, 61] and asthma[61].
24
25 538 The technology's multiple advantages of ultra-sensitivity, portability, online sampling and short
26
27 539 analysis time (typical analysis time of 10 minutes) with real time detection, brings a promise to
28
29 540 provide immediate and potentially reliable results for point of care breath diagnostics. Another
30
31 541 concept with IMS devices is that once the required breath signatures have been discovered using GC-
32
33 542 MS, IMS offers the potential to be 'tuned' for selective detection of VOC.

34
35
36 543 **3.5. Atmospheric pressure chemical ionisation- mass spectrometry (APCI-MS) semiportable**
37
38 544 **compact version:** is one of less sensitive but more affordable versions of mass spectrometers released
39
40 545 to the commercial market in recent years. The device uses APCI to produce ions. Although the most
41
42 546 common use of APCI-MS systems is the detection in liquid chromatography applications, the
43
44 547 technique has proven to be a valuable tool for direct measurement of VOC in air[62, 63] food[64, 65]
45
46 548 and breath[66, 67]. Recently, the technique has shown potential for online, real time profiling of
47
48 549 pseudo-metabolites in exhaled breath [68] with sensitivity comparable with other techniques. By
49
50 550 combining miniaturised MS technology with APCI techniques, adequate quality of on-site, real time
51
52 551 measurements with minimal or no sample preparation requirement can be provided. This is a desirable
53
54 552 outcome as it overcomes main limitation of using standard breath analysis method in clinical setting,
55
56
57
58
59
60

1
2
3 553 which is a need for breath sample collection followed by desorption and time-consuming laboratory
4
5 554 analysis.

6
7 587 There remains an overall lack of standardisation and rigour across these technologies which hindered
8
9 588 previous advancements in breath discovery; something we intend to minimize.

11 12 589 **4. Chemometric processing and data analysis:**

13
14 590 GC-MS breath data will be aligned, deconvoluted and the features for each participant will be
15
16 591 extracted. The extracted features will be grouped and classified by retention index and mass spectrum.
17
18 592 The registered and aligned data will be linked to participant meta-data to generate a breath matrix.

19
20
21 593 The breath matrix is a $n \times p$ matrix where n is the number of subjects and p is the number of VOC. The
22
23 594 breath matrix is high dimensional with $p \gg n$ and many potentially correlated VOC. In view of this, we
24
25 595 will employ sparse partial least squares discriminant analysis (sPLS-DA)[69] to investigate which of
26
27 596 the VOC can identify breathlessness. We will also investigate which of the VOC can discriminate
28
29 597 between the different disease states including acute exacerbations of asthma and COPD and
30
31 598 Pneumonia. In addition to the supervised methods, unsupervised methods will be explored,
32
33 599 specifically sparse principle component analysis (sPCA)[70].

34
35
36 600 Extracted VOC will also be investigated. Relationships between VOC and patient reported acute
37
38 601 breathlessness will be analysed using logistic regression model. VOC associated with patient
39
40 602 associated acute breathlessness will be incorporated into multinomial logistic regression models in
41
42 603 conjunction with CRP, BNP, blood eosinophils and Troponin-I, pathology biomarkers currently in use
43
44 604 for diagnosing undifferentiated breathlessness. In addition to the conventional binary and multinomial
45
46 605 logistic regression models, regression models [71].

47 48 49 606 **5. Ethics and dissemination:**

50
51
52 607 The study has obtained full ethical approval from the London South East Research ethics Committee,
53
54 608 REC reference 16/LO/1747. IRAS project ID 198921. Publications will be prepared according to the
55
56 609 MRC-EMBER consortium agreement and the University of Leicester publications policy. All

1
2
3 610 intended publications will be submitted to the EMBER executive board for review and comments
4
5 611 within 60 days of journal submission. Authorship will be according to contribution and internationally
6
7 612 recognised guidance on journal authorship.
8
9

10 613 **6. Study dates:** 01/2/2017 – 30/10/2020

11
12 614 **7. Authors' contributions:**

13
14 615 WI and SS drafted the manuscript and all co-authors critically revised and contributed to the
15
16 616 manuscript. All co-authors contributed to the study design and development. SS is the Chief
17
18 617 investigator for the acute VOC study.
19

20 618

21
22 619 **Protocol version:** Version 4, 1st April 2018
23

24 620

25
26 621 **Funding statement:**

27
28
29 622 This study has been funded by the Medical research Council (MRC) and Engineering and Physical
30
31 623 Sciences Research Council (EPSRC) and is sponsored by the University of Leicester. The children's
32
33 624 component of this study is also being supported by The Midlands Asthma and Allergy Research
34
35 625 Association (MAARA). Grant number [MR/N005880/1]
36

37
38 626 **Competing interests:** SS has performed advisory services for Owlstone Medical.
39

40 627

41 628

42 629

43 630

44 631

45 632

46 633

47 634

48 635

49 636

1
2
3 637
4 638
5
6 639
7
8 640
9
10 641

11 **Figure legends:**

12 **Figure (1)** Relationship between lung proximity and degree of invasiveness of different lung
13 matrices. The figure plots the level of invasiveness of various lung matrices in relation to their
14 proximity to the lung. Given their pathological relevance, the degree of invasiveness of
15 bronchoalveolar lavage (BAL) and lung biopsy makes them less favourable in diagnosing respiratory
16 diseases.
17
18
19
20
21
22

23
24 **Figure (2)** Multi-instrument use in breath sampling. Figure illustrates the various combinations of
25 offline and online devices used in breath sampling and the relevant pros and cons. Offline and online
26 technologies are used for the discovery and validation phases of the study respectively.
27
28
29

30
31 **Figure (3)** Study flow chart. Figure outlines the patient journey from admission through to discharge
32 and follow up. Participants with self-reported acute breathlessness presenting to University Hospitals
33 of Leicester are recruited within 24 hours following a senior decision maker review. Breath sampling
34 is carried out on the first visit, at recruitment, and the second visit, up to 6 months after discharge.
35 Patients are admitted through the standard operational emergency medical streaming and care
36 pathways at the University Hospitals of Leicester NHS Trust. Early outcomes (hospital readmission)
37 are measured at 30 and 60 days and late outcomes (mortality) including respiratory and all-cause
38 mortality are measured at 2 years. Assessments carried out at each time point are summarised in
39
40
41
42
43
44
45
46

47 **Table 1.**

48
49 **Figure (4)** Multi-instrument use in breath sampling. Operational space of the analytical technologies
50 used in EMBER for the analysis of volatile organic compounds (VOC) in exhaled breath, including
51 proton transfer reaction-mass spectrometry (PTR-MS), atmospheric pressure chemical ionisation-
52 mass spectrometry (CMS), gas chromatography-ion mobility spectrometry (GC-IMS), gas
53
54
55
56
57
58
59
60

1
2
3 664 chromatography-mass spectrometry (GC-MS) and two-dimensional gas chromatography-mass
4
5 665 spectrometry (GCxGC). Comparing the typical molar mass range detectable; selectivity in detection
6
7 666 owing to the type of ionisation involved and the proton affinity of analytes; and the inclusion of a
8
9 667 chromatographic separation affecting total time of analysis. The online technologies involving
10
11 668 chemical ionisation (PTR-MS, CMS and GC-IMS) can be used in-clinic owing to short analysis
12
13 669 times, but only detect lower molar mass molecules with a proton affinity higher than 697 KJ mol⁻¹.
14
15 670 Offline chromatographic techniques (GC-MS and GCxGC) detect a wider range of compounds
16
17 671 independent of proton affinity; however, the techniques have longer analysis times and involve
18
19 672 sample transportation and storage.
20
21 673
22
23 674
24
25 675
26
27 676
28
29 677
30
31 678
32
33 679
34
35 680
36
37 681
38
39 682
40
41 683
42
43 684
44
45 685
46
47 686
48
49 687
50
51 688
52
53 689
54
55 690
56
57
58
59
60

691

692

693

694

695 **8. References**

- 696 1. Stephenson, J.J., et al., *Clinical and economic burden of dyspnea and other COPD symptoms*
697 *in a managed care setting*. Int J Chron Obstruct Pulmon Dis, 2017. **12**: p. 1947-1959.
- 698 2. David Ponka, M.K., *Top differential diagnoses in family medicine Dyspnoea Can Fam*
699 *Physician*. Canadian Family Physician, 2008.
- 700 3. Figarska, S.M., H.M. Boezen, and J.M. Vonk, *Dyspnea severity, changes in dyspnea status and*
701 *mortality in the general population: the Vlagtwedde/Vlaardingen study*. European Journal of
702 Epidemiology, 2012. **27**(11): p. 867-876.
- 703 4. Parshall, M.B., et al., *An Official American Thoracic Society Statement: Update on the*
704 *Mechanisms, Assessment, and Management of Dyspnea*. American Journal of Respiratory
705 and Critical Care Medicine, 2012. **185**(4): p. 435-452.
- 706 5. Lim, W.S., et al., *BTS guidelines for the management of community acquired pneumonia in*
707 *adults: update 2009*. Thorax, 2009. **64 Suppl 3**: p. iii1-55.
- 708 6. Leuzzi, G., et al., *C-reactive protein level predicts mortality in COPD: a systematic review and*
709 *meta-analysis*. Eur Respir Rev, 2017. **26**(143).
- 710 7. Moe, G.W., *B-type natriuretic peptide in heart failure*. Curr Opin Cardiol, 2006. **21**(3): p. 208-
711 14.
- 712 8. Ponikowski, P., et al., *2016 ESC Guidelines for the diagnosis and treatment of acute and*
713 *chronic heart failure: The Task Force for the diagnosis and treatment of acute and chronic*
714 *heart failure of the European Society of Cardiology (ESC) Developed with the special*
715 *contribution of the Heart Failure Association (HFA) of the ESC*. Eur Heart J, 2016. **37**(27): p.
716 2129-2200.
- 717 9. Alex Sakula, *Charcot leyden crystals and crushmann spirals*. Thorax, 1986.
- 718 10. Calhoun, W.J., J. Sedgwick, and W.W. Busse, *The role of eosinophils in the pathophysiology of*
719 *asthma*. Ann N Y Acad Sci, 1991. **629**: p. 62-72.
- 720 11. Pavord, I.D., *Eosinophilic phenotypes of airway disease*. Ann Am Thorac Soc, 2013. **10 Suppl**:
721 p. S143-9.
- 722 12. Tashkin, D.P. and M.E. Wechsler, *Role of eosinophils in airway inflammation of chronic*
723 *obstructive pulmonary disease*. International Journal of Chronic Obstructive Pulmonary
724 Disease, 2018. **13**: p. 335-349.
- 725 13. Bafadhel, M., et al., *Blood Eosinophils to Direct Corticosteroid Treatment of Exacerbations of*
726 *Chronic Obstructive Pulmonary Disease: A Randomized Placebo-Controlled Trial*. American
727 Journal of Respiratory and Critical Care Medicine, 2012. **186**(1): p. 48-55.
- 728 14. Bedossa, P. and K. Patel, *Biopsy and Noninvasive Methods to Assess Progression of*
729 *Nonalcoholic Fatty Liver Disease*. Gastroenterology, 2016. **150**(8): p. 1811-1822 e4.
- 730 15. Fens, N., et al., *Exhaled breath analysis by electronic nose in airways disease. Established*
731 *issues and key questions*. Clin Exp Allergy, 2013. **43**(7): p. 705-15.
- 732 16. Probert, C.S., et al., *Volatile organic compounds as diagnostic biomarkers in gastrointestinal*
733 *and liver diseases*. J Gastrointestin Liver Dis, 2009. **18**(3): p. 337-43.
- 734 17. Schubert, J.K., et al., *Breath analysis in critically ill patients: potential and limitations*. Expert
735 Rev Mol Diagn, 2004. **4**(5): p. 619-29.

- 1
2
3 736 18. Fowler, S.J., et al., *Surveillance for lower airway pathogens in mechanically ventilated*
4 737 *patients by metabolomic analysis of exhaled breath: a case-control study*. Thorax, 2015.
5 738 **70**(4): p. 320-5.
- 6 739 19. Florian Gahleitner, C.G.-H., Caroline S Beardsmore, Hitesh C Pandya, CL Paul Thomas,
7 740 *Metabolomics pilot study to identify VOC of childhood asthma in exhaled breath*. Bioanalysis,
8 741 2013.
- 9 742 20. Christiansen, A., et al., *A systematic review of breath analysis and detection of volatile*
10 743 *organic compounds in COPD*. J Breath Res, 2016. **10**(3): p. 034002.
- 11 744 21. de Vries, R., et al., *Integration of electronic nose technology with spirometry: validation of a*
12 745 *new approach for exhaled breath analysis*. J Breath Res, 2015. **9**(4): p. 046001.
- 13 746 22. Shafiek, H., et al., *Using the Electronic Nose to Identify Airway Infection during COPD*
14 747 *Exacerbations*. PLoS One, 2015. **10**(9): p. e0135199.
- 15 748 23. Finamore, P., et al., *Analysis of volatile organic compounds: an innovative approach to heart*
16 749 *failure characterization in older patients*. J Breath Res, 2018. **12**(2): p. 026007.
- 17 750 24. MRC Molecular Pathology, Review 20 Aug.
- 18 751 25. Medical Research Council, *Molecular Pathology Nodes*. 2014: MRC website.
- 19 752 26. Corradi, M., et al., *Aldehydes in exhaled breath condensate of patients with chronic*
20 753 *obstructive pulmonary disease*. Am J Respir Crit Care Med, 2003. **167**(10): p. 1380-6.
- 21 754 27. Ibrahim, B., et al., *Non-invasive phenotyping using exhaled volatile organic compounds in*
22 755 *asthma*. Thorax, 2011. **66**(9): p. 804-9.
- 23 756 28. Larstad, M., et al., *Determination of malondialdehyde in breath condensate by high-*
24 757 *performance liquid chromatography with fluorescence detection*. J Chromatogr B Analyt
25 758 Technol Biomed Life Sci, 2002. **766**(1): p. 107-14.
- 26 759 29. F. Y. Hsieh, D.A.B., Michael D. Larsen,, *A simple method of sample size calculation for linear*
27 760 *and logistic regression* Statis Med, 1998.
- 28 761 30. Hajian-Tilaki, K., *Sample size estimation in diagnostic test studies of biomedical informatics*. J
29 762 Biomed Inform, 2014. **48**: p. 193-204.
- 30 763 31. Quan, H., et al., *Updating and validating the Charlson comorbidity index and score for risk*
31 764 *adjustment in hospital discharge abstracts using data from 6 countries*. Am J Epidemiol,
32 765 2011. **173**(6): p. 676-82.
- 33 766 32. Roffman, C.E., J. Buchanan, and G.T. Allison, *Charlson Comorbidities Index*. J Physiother,
34 767 2016. **62**(3): p. 171.
- 35 768 33. Bjarnason, H. and P. Hougaard, *Fisher information for two gamma frailty bivariate Weibull*
36 769 *models*. Lifetime Data Anal, 2000. **6**(1): p. 59-71.
- 37 770 34. Lawrence Gould, A., et al., *Joint modeling of survival and longitudinal non-survival data:*
38 771 *current methods and issues. Report of the DIA Bayesian joint modeling working group*. Stat
39 772 Med, 2015. **34**(14): p. 2181-95.
- 40 773 35. Steer, J., et al., *Dyspnoea severity and pneumonia as predictors of in-hospital mortality and*
41 774 *early readmission in acute exacerbations of COPD*. Thorax, 2012. **67**(2): p. 117-21.
- 42 775 36. Tsimogianni, A.M., et al., *Predictors of Outcome After Exacerbation of Chronic Obstructive*
43 776 *Pulmonary Disease*. Journal of General Internal Medicine, 2009. **24**(9): p. 1043-1048.
- 44 777 37. Kitchen, S., et al., *LATE-BREAKING ABSTRACT: Breathe free: Open source development of a*
45 778 *breath sampler by a consortium of breath researchers*. European Respiratory Journal, 2015.
46 779 **46**(suppl 59).
- 47 780 38. Oostveen, E., et al., *The forced oscillation technique in clinical practice: methodology,*
48 781 *recommendations and future developments*. Eur Respir J, 2003. **22**(6): p. 1026-41.
- 49 782 39. Pendharkar, S. and S. Mehta, *The clinical significance of exhaled nitric oxide in asthma*.
50 783 Canadian Respiratory Journal : Journal of the Canadian Thoracic Society, 2008. **15**(2): p. 99-
51 784 106.
- 52 785 40. Yates, D.H., *Role of exhaled nitric oxide in asthma*. Immunol Cell Biol, 2001. **79**(2): p. 178-90.

- 1
2
3 786 41. Lang, R.M., et al., *Recommendations for chamber quantification: a report from the American*
4 787 *Society of Echocardiography's Guidelines and Standards Committee and the Chamber*
5 788 *Quantification Writing Group, developed in conjunction with the European Association of*
6 789 *Echocardiography, a branch of the European Society of Cardiology.* J Am Soc Echocardiogr,
7 790 2005. **18**(12): p. 1440-63.
- 8 791 42. J. Throck Watson, O.D.S., *Introduction to Mass Spectrometry: Instrumentation, Applications,*
9 792 *and Strategies for Data Interpretation, 4th Edition.* 2007: Wiley. 862.
- 10 793 43. Buszewski, B., et al., *Human exhaled air analytics: biomarkers of diseases.* Biomed
11 794 Chromatogr, 2007. **21**(6): p. 553-66.
- 12 795 44. Amann, A., et al., *Applications of breath gas analysis in medicine.* International Journal of
13 796 Mass Spectrometry, 2004. **239**(2-3): p. 227-233.
- 14 797 45. Almstetter, M.F., P.J. Oefner, and K. Dettmer, *Comprehensive two-dimensional gas*
15 798 *chromatography in metabolomics.* Anal Bioanal Chem, 2012. **402**(6): p. 1993-2013.
- 16 799 46. Das, M.K., et al., *Investigation of Gender-Specific Exhaled Breath Volatome in Humans by*
17 800 *GCxGC-TOF-MS.* Analytical Chemistry, 2014. **86**(2): p. 1229-1237.
- 18 801 47. Phillips, M., et al., *Detection of an extended human volatome with comprehensive two-*
19 802 *dimensional gas chromatography time-of-flight mass spectrometry.* PLoS One, 2013. **8**(9): p.
20 803 e75274.
- 21 804 48. Gruber, B., et al., *Vacuum ultraviolet absorption spectroscopy in combination with*
22 805 *comprehensive two-dimensional gas chromatography for the monitoring of volatile organic*
23 806 *compounds in breath gas: A feasibility study.* J Chromatogr A, 2016. **1464**: p. 141-6.
- 24 807 49. Gruber, B., et al., *Breath gas monitoring during a glucose challenge by a combined PTR-*
25 808 *QMS/GCxGC-TOFMS approach for the verification of potential volatile biomarkers.* J Breath
26 809 Res, 2016. **10**(3): p. 036003.
- 27 810 50. Mellors, T.R., et al., *A new method to evaluate macaque health using exhaled breath: A case*
28 811 *study of M. tuberculosis in a BSL-3 setting.* J Appl Physiol (1985), 2017. **122**(3): p. 695-701.
- 29 812 51. Phillips, M., et al., *Breath biomarkers of whole-body gamma irradiation in the Gottingen*
30 813 *minipig.* Health Phys, 2015. **108**(5): p. 538-46.
- 31 814 52. Bajtarevic, A., et al., *Noninvasive detection of lung cancer by analysis of exhaled breath.* BMC
32 815 Cancer, 2009. **9**: p. 348.
- 33 816 53. Lavra, L., et al., *Investigation of VOCs associated with different characteristics of breast*
34 817 *cancer cells.* Sci Rep, 2015. **5**: p. 13246.
- 35 818 54. Zhou, W., et al., *Exhaled breath online measurement for cervical cancer patients and healthy*
36 819 *subjects by proton transfer reaction mass spectrometry.* Anal Bioanal Chem, 2017. **409**(23):
37 820 p. 5603-5612.
- 38 821 55. Fernández del Río, R., et al., *Volatile Biomarkers in Breath Associated With Liver Cirrhosis —*
39 822 *Comparisons of Pre- and Post-liver Transplant Breath Samples.* EBioMedicine, 2015. **2**(9): p.
40 823 1243-1250.
- 41 824 56. O'Hara, M.E., et al., *Limonene in exhaled breath is elevated in hepatic encephalopathy.* J
42 825 Breath Res, 2016. **10**(4): p. 046010.
- 43 826 57. Tomasz Ligor, C.A., Anton Amann, Konrad Schwarz, Bogusaw Buszewski, Wojciech ebrowski,,
44 827 *<Comparison of Proton Transfer Reaction-Mass Spectrometry and Gas Chromatography-*
45 828 *Mass Spectrometry in Analysis of Breath Samples.pdf>.* Chem. Anal., 2009.
- 46 829 58. Darwiche, K., et al., *Bronchoscopically obtained volatile biomarkers in lung cancer.* Lung,
47 830 2011. **189**(6): p. 445-52.
- 48 831 59. Handa, H., et al., *Exhaled Breath Analysis for Lung Cancer Detection Using Ion Mobility*
49 832 *Spectrometry.* PLoS ONE, 2014. **9**(12): p. e114555.
- 50 833 60. Bessa, V., et al., *Detection of volatile organic compounds (VOCs) in exhaled breath of patients*
51 834 *with chronic obstructive pulmonary disease (COPD) by ion mobility spectrometry.*
52 835 International Journal for Ion Mobility Spectrometry, 2011. **14**(1): p. 7-13.

- 1
2
3 836 61. Schivo, M., et al., *Mobile instrumentation platform to distinguish airway disorders*. Journal of
4 837 breath research, 2013. **7**(1): p. 017113-017113.
- 5 838 62. Badjagbo, K., et al., *Direct atmospheric pressure chemical ionization-tandem mass*
6 839 *spectrometry for the continuous real-time trace analysis of benzene, toluene, ethylbenzene,*
7 840 *and xylenes in ambient air*. J Am Soc Mass Spectrom, 2009. **20**(5): p. 829-36.
- 8 841 63. Healy, R.M., et al., *A multi-year study of VOC emissions at a chemical waste disposal facility*
9 842 *using mobile APCI-MS and LPCI-MS instruments*. Environ Pollut, 2018. **232**: p. 220-228.
- 10 843 64. Linforth, R. and A.J. Taylor, *Persistence of Volatile Compounds in the Breath after Their*
11 844 *Consumption in Aqueous Solutions*. Journal of Agricultural and Food Chemistry, 2000. **48**(11):
12 845 p. 5419-5423.
- 13 846 65. Perez-Hurtado, P., et al., *Direct analysis of volatile organic compounds in foods by headspace*
14 847 *extraction atmospheric pressure chemical ionisation mass spectrometry*. Rapid Commun
15 848 Mass Spectrom, 2017. **31**(22): p. 1947-1956.
- 16 849 66. Berchtold, C., L. Meier, and R. Zenobi, *Evaluation of extractive electrospray ionization and*
17 850 *atmospheric pressure chemical ionization for the detection of narcotics in breath*.
18 851 International Journal of Mass Spectrometry, 2011. **299**(2-3): p. 145-150.
- 19 852 67. Kapishon, V., et al., *Atmospheric pressure chemical ionization mass spectrometry of pyridine*
20 853 *and isoprene: potential breath exposure and disease biomarkers*. J Breath Res, 2013. **7**(2): p.
21 854 026005.
- 22 855 68. Heaney, L.M., et al., *Real-time monitoring of exhaled volatiles using atmospheric pressure*
23 856 *chemical ionization on a compact mass spectrometer*. Bioanalysis, 2016. **8**(13): p. 1325-36.
- 24 857 69. Lê Cao, K.-A., S. Boitard, and P. Besse, *Sparse PLS discriminant analysis: biologically relevant*
25 858 *feature selection and graphical displays for multiclass problems*. BMC Bioinformatics, 2011.
26 859 **12**: p. 253-253.
- 27 860 70. Zou, H., T. Hastie, and R. Tibshirani, *Sparse Principal Component Analysis*. Journal of
28 861 Computational and Graphical Statistics, 2006. **15**(2): p. 265-286.
- 29 862 71. Zou H, H.T., *Regularization and Variable Selection via the Elastic Net*. Journal of the Royal
30 863 Statistical Society, 2005.
- 31
32
33 864
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

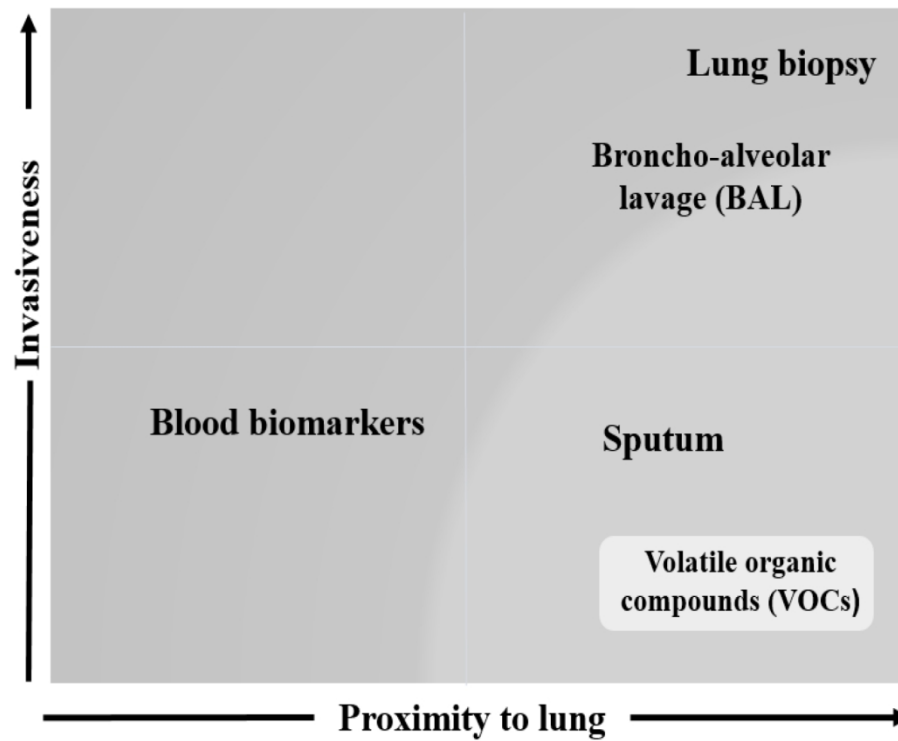


Figure (1): Relationship between lung proximity and degree of invasiveness of different lung matrices. The figure plots the level of invasiveness of various lung matrices in relation to their proximity to the lung. Given their pathological relevance, the degree of invasiveness of bronchoalveolar lavage (BAL) and lung biopsy makes them less favourable in diagnosing respiratory diseases.

144x115mm (300 x 300 DPI)

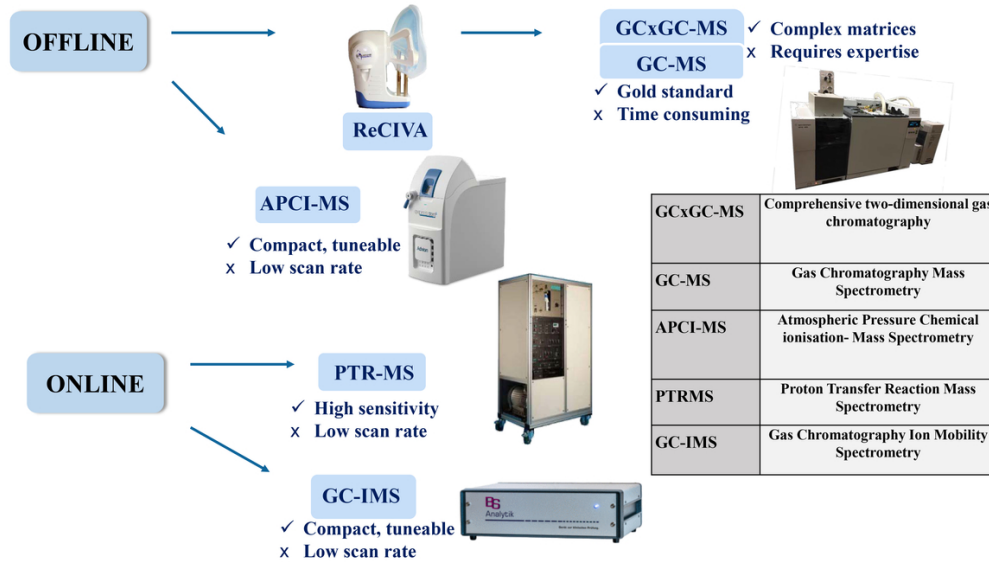
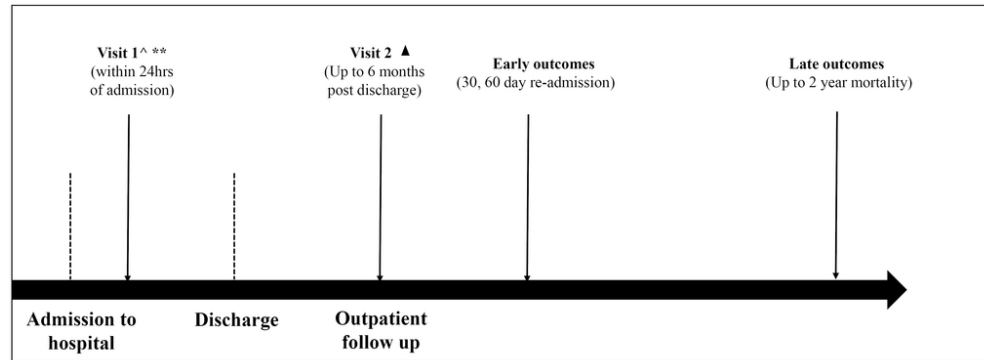


Figure (2): Multi-instrument use in breath sampling. Figure illustrates the various combinations of offline and online devices used in breath sampling and the relevant pros and cons. Offline and online technologies are used for the discovery and validation phases of the study respectively.

97x55mm (300 x 300 DPI)



- ^ Following senior decision maker review
 ** Breath sampling (ReCIVA-GC-MS, ReCIVA- GCxGC-MS, GC-IMS, APCI-MS)
 ▲ Breath sampling (ReCIVA-GC-MS, ReCIVA- GCxGC-MS, PTR-MS, APCI-MS)
 ---- Following University Hospitals of Leicester streaming and clinical care pathways

Figure (3): Study flow chart. Figure outlines the patient journey from admission through to discharge and follow up. Participants with self-reported acute breathlessness presenting to University Hospitals of Leicester are recruited within 24 hours following a senior decision maker review. Breath sampling is carried out on the first visit, at recruitment, and the second visit, up to 6 months after discharge. Patients are admitted through the standard operational emergency medical streaming and care pathways at the University Hospitals of Leicester NHS Trust. Early outcomes (hospital readmission) are measured at 30 and 60 days and late outcomes (mortality) including respiratory and all-cause mortality are measured at 2 years. Assessments carried out at each time point are summarised in Table 1.

86x43mm (300 x 300 DPI)

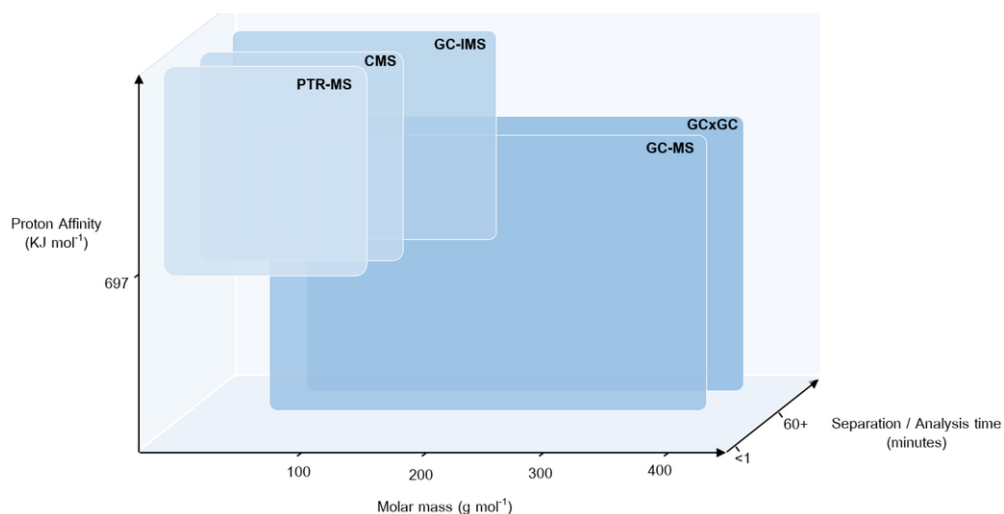


Figure (4): Multi-instrument use in breath sampling. Operational space of the analytical technologies used in EMBER for the analysis of volatile organic compounds (VOC) in exhaled breath, including proton transfer reaction-mass spectrometry (PTR-MS), atmospheric pressure chemical ionisation-mass spectrometry (CMS), gas chromatography-ion mobility spectrometry (GC-IMS), gas chromatography-mass spectrometry (GC-MS) and two-dimensional gas chromatography-mass spectrometry (GCxGC). Comparing the typical molar mass range detectable; selectivity in detection owing to the type of ionisation involved and the proton affinity of analytes; and the inclusion of a chromatographic separation affecting total time of analysis. The online technologies involving chemical ionisation (PTR-MS, CMS and GC-IMS) can be used in-clinic owing to short analysis times, but only detect lower molar mass molecules with a proton affinity higher than 697 KJ mol⁻¹. Offline chromatographic techniques (GC-MS and GCxGC) detect a wider range of compounds independent of proton affinity; however, the techniques have longer analysis times and involve sample transportation and storage.

88x44mm (300 x 300 DPI)

BMJ Open

Assessment of Breath Volatile Organic Compounds in Acute Cardio-respiratory Breathlessness: A protocol describing a Prospective Real World Observational Study

Journal:	<i>BMJ Open</i>
Manuscript ID	bmjopen-2018-025486.R1
Article Type:	Protocol
Date Submitted by the Author:	11-Oct-2018
Complete List of Authors:	<p>Ibrahim, Wadah; Leicester NIHR Biomedical Research Centre (Respiratory theme); College of Life Sciences, Department of Infection, Immunity and Inflammation, University of Leicester</p> <p>Wilde, Michael; University of Leicester, Department of Chemistry</p> <p>Cordell, Rebecca; University of Leicester, Department of Chemistry</p> <p>Salman, Dahlia; Loughborough University, Department of Chemistry</p> <p>Ruszkiewicz, Dorota; Loughborough University, Department of Chemistry</p> <p>Bryant, Luke; University of Leicester, Department of Chemistry</p> <p>Richardson, Matthew; Leicester NIHR Biomedical Research Centre (Respiratory theme); University of Leicester, Infection, Inflammation and Immunity</p> <p>Free, Robert; Leicester NIHR Biomedical Research Centre (Respiratory theme); University of Leicester, Infection, Inflammation and Immunity</p> <p>Zhao, Bo; Leicester NIHR Biomedical Research Centre (Respiratory theme); University of Leicester, Infection, Inflammation and Immunity</p> <p>Yousuf, Ahmed; University of Leicester, Infection, Inflammation and Immunity; Leicester NIHR Biomedical Research Centre (Respiratory theme)</p> <p>White, Christobelle; University of Leicester, Infection, Inflammation and Immunity; Leicester NIHR Biomedical Research Centre (Respiratory theme)</p> <p>Russell, Richard; University of Leicester, Infection, Inflammation and Immunity; Leicester NIHR Biomedical Research Centre (Respiratory theme)</p> <p>Jones, Sheila; University Hospitals of Leicester NHS Trust</p> <p>Patel, Bharti; University Hospitals of Leicester NHS Trust</p> <p>Awal, Asia; University Hospitals of Leicester NHS Trust</p> <p>Phillips, Rachael; NIHR Leicester Clinical Research Facility</p> <p>Fowkes, Graham; NIHR Leicester Clinical Research Facility</p> <p>McNally, Teresa; Paediatric Clinical Investigation Centre</p> <p>Foxon, Clare; Paediatric Clinical Investigation Centre</p> <p>Bhatt, Hetan; Leicester NIHR Biomedical Research Centre (Respiratory theme)</p> <p>Peltrini, Rosa; Leicester NIHR Biomedical Research Centre (Respiratory theme); University of Leicester, Department Infection, Inflammation and Immunity</p> <p>Singapuri, Amisha; Leicester NIHR Biomedical Research Centre (Respiratory theme)</p> <p>Hargadon, Beverley; Leicester NIHR Biomedical Research Centre</p>

	(Respiratory theme); University of Leicester, Infection, Inflammation and Immunity Suzuki, T; Department of Cardiovascular Sciences, University of Leicester, , Cardiovascular Research Centre; Leicester NIHR Biomedical Research Centre (Cardiovascular theme) Ng, Leong; Department of Cardiovascular Sciences, University of Leicester, Cardiovascular Research Centre; Leicester NIHR Biomedical Research Centre (Cardiovascular theme) Gaillard, Erol; Leicester NIHR Biomedical Research Centre (Respiratory theme); Paediatric Clinical Investigation Centre Beardsmore, Caroline; Leicester NIHR Biomedical Research Centre (Respiratory theme); Paediatric Clinical Investigation Centre Ryanna, Kimuli; University Hospitals of Leicester NHS Trust Pandya, Hitesh; Discovery Medicine, Respiratory Therapeutic Area, GlaxoSmithKline PLC Coates, Tim; University of Leicester, Infection, Inflammation and Immunity; Department of Cardiovascular Sciences, University of Leicester Monks, Paul; University of Leicester, Infection, Inflammation and Immunity; University of Leicester, Department of Chemistry Greening, Neil; Leicester NIHR Biomedical Research Centre (Respiratory theme); University of Leicester, Infection, Inflammation and Immunity Brightling, Christopher; Leicester NIHR Biomedical Research Centre (Respiratory theme); University of Leicester, Infection, Inflammation and Immunity Thomas, Paul; Loughborough University, Department of Chemistry Siddiqui, Salman; Leicester NIHR Biomedical Research Centre (Respiratory theme); University of Leicester, Infection, Inflammation and Immunity
Primary Subject Heading:	Respiratory medicine
Secondary Subject Heading:	Research methods, Respiratory medicine
Keywords:	Breathlessness, Observational study, Volatile Organic Compound, Breath analysis

SCHOLARONE™
Manuscripts

1 Assessment of Breath Volatile Organic Compounds in Acute Cardio- 2 respiratory Breathlessness: A protocol describing a Prospective Real World 3 Observational Study

4
5 Wadah Ibrahim^{1,2}, Michael Wilde³, Rebecca Cordell³, Dahlia Salman⁴, Dorota Ruzskiewicz⁴, Luke
6 Bryant³, Matthew Richardson^{1,2}, Robert C. Free^{1,2}, Bo Zhao^{1,2}, Ahmed Yousuf^{1,2}, Christobelle White^{1,2},
7 Richard J Russell^{1,2}, Sheila Jones⁵, Bharti Patel⁵, Asia Awal⁵, Rachael Phillips⁶, Graham Fowkes⁶, Teresa
8 McNally⁷, Clare Foxon⁷, Hetan Bhatt^{1,2}, Rosa Peltrini^{1,2}, Amisha Singapuri^{1,2}, Beverley Hargadon^{1,2},
9 Toru Suzuki^{8,9}, Leong L. Ng^{8,9}, Erol A Gaillard^{2,7}, Caroline Beardsmore^{2,7}, Kimuli Ryanna⁵, Hitesh
10 Pandya¹¹, Tim Coates^{2,10}, Paul S. Monks^{2,3}, Neil Greening^{1,2}, Chris E Brightling^{1,2}, Paul Thomas⁴, Salman
11 Siddiqui^{1,2}

12
13
14 wadah.ibrahim@leicester.ac.uk; mjw77@leicester.ac.uk; rc145@leicester.ac.uk; d.salman@lboro.ac.uk;
15 d.m.ruzskiewicz2@lboro.ac.uk; lb368@leicester.ac.uk; mr251@leicester.ac.uk; rcf8@leicester.ac.uk;
16 bo.zhao@leicester.ac.uk; ajy7@leicester.ac.uk; cmjw3@leicester.ac.uk; rjr22@leicester.ac.uk;
17 rjr22@leicester.ac.uk; sheila.jones@uhl-tr.nhs.uk; bharti.m.patel@uhl-tr.nhs.uk; asia.awal@uhl-tr.nhs.uk;
18 rachael.phillips@uhl-tr.nhs.uk; graham.fowkes@uhl-tr.nhs.uk; tm23@leicester.ac.uk; clare.foxon@uhl-tr.nhs.uk;
19 hetan.bhatt@uhl-tr.nhs.uk; rp417@leicester.ac.uk; amisha.singapuri@uhl-tr.nhs.uk; beverley.hargadon@uhl-
20 tr.nhs.uk; ts263@le.ac.uk; lln1@le.ac.uk; eag15@leicester.ac.uk; csb@leicester.ac.uk; kimuli.ryanna@uhl-
21 tr.nhs.uk; hitesh.x.pandya@gsk.com; tim.coates@leicester.ac.uk; paul.monks@leicester.ac.uk;
22 neil.greening@leicester.ac.uk; ceb17@leicester.ac.uk; C.L.P.Thomas@lboro.ac.uk; ss338@leicester.ac.uk

23
24 Corresponding author: Professor Salman Siddiqui, ss338@leicester.ac.uk

25 NIHR Leicester Biomedical Research Centre (Respiratory Theme)

26 Glenfield Hospital,

27 Groby Road

28 Leicester

29 LE3 9QP

30 t: +44 116 256 3841

31 e: ss338@leicester.ac.uk

32 33 Authors' affiliations:

34 ¹Leicester NIHR Biomedical Research Centre (Respiratory theme), Glenfield Hospital, Groby Road, Leicester LE3

35 9QP

1
2
3 36 ² College of Life Sciences, Department of Infection, Immunity and Inflammation, University of Leicester,
4
5 37 University Road, Leicester LE1 7RH

6
7 38 ³ Department of Chemistry, University of Leicester, University Road, Leicester, LE1 7RH

8
9 39 ⁴ Department of Chemistry, Loughborough University, Epinal Way, Loughborough, LE11 3TU

10
11 40 ⁵ University Hospitals of Leicester NHS Trust, Glenfield Hospital, Groby Road, Leicester LE3 9QP

12
13 41 ⁶ NIHR Leicester Clinical Research Facility, Research Space, Level 0, Balmoral Building, Leicester Royal
14
15 42 Infirmary, Infirmary Square, Leicester, LE1 5WW

16
17 43 ⁷ Paediatric Clinical Investigation Centre, Leicester Respiratory Biomedical Research Centre, University of
18
19 44 Leicester, Leicester Royal Infirmary, Leicester, LE2 7LX

20
21 45 ⁸ Department of Cardiovascular Sciences, University of Leicester, Cardiovascular Research Centre, Glenfield
22
23 46 General Hospital, Leicester LE3 9QP.

24
25 47 ⁹ Leicester NIHR Biomedical Research Centre (Cardiovascular theme), Glenfield Hospital, Groby Road, Leicester
26
27 48 LE3 9QP

28
29 49 ¹⁰ Department of Cardiovascular Sciences, University of Leicester, Level G, Jarvis building RMO, Infirmary Square,
30
31 50 Leicester LE1 5WW

32
33 51 ¹¹ Discovery Medicine, Respiratory Therapeutic Area, GlaxoSmithKline PLC, Gunnels Wood Road, Stevenage,
34
35 52 Hertfordshire, SG1 2NY

36
37 53

38
39 54

40
41 55

42
43 56

44
45 57

46
47 58

48
49 59

50
51 60

52
53 61

54
55 62

56
57 63

1
2
3 644
5 656
7 668
9 6710
11 6812
13
14 69 **Abstract**15
16 70 **Introduction**

17
18 71 Patients presenting with acute undifferentiated breathlessness are commonly encountered in
19
20 72 admissions units across the United Kingdom. Existing blood biomarkers have clinical utility in
21
22 73 distinguishing patients with single organ pathologies but have poor discriminatory power in multi
23
24 74 factorial presentations. Evaluation of volatile organic compounds (VOC) in exhaled breath offers the
25
26 75 potential to develop biomarkers of disease states that underpin acute cardio-respiratory breathlessness,
27
28 76 owing to their proximity to the cardio-respiratory system. To date there has been no systematic evaluation
29
30 77 of VOC in acute cardio-respiratory breathlessness. The proposed study will seek to use both offline and
31
32 78 online VOC technologies to evaluate the predictive value of VOC in identifying common conditions that
33
34 79 present with acute cardio-respiratory breathlessness.

35
36
37 80 **Methods and analysis**

38
39
40 81 A prospective real world observational study carried out across three acute admissions units within
41
42 82 Leicestershire. Participants with self-reported acute breathlessness, with a confirmed primary diagnosis of
43
44 83 either acute heart failure, community acquired pneumonia and acute exacerbation of asthma or COPD
45
46 84 will be recruited within 24 hours of admission. Additionally, school age children admitted with severe
47
48 85 asthma will be evaluated. All participants will undergo breath sampling on admission and upon recovery
49
50 86 following discharge. A range of online technologies including: proton-transfer-reaction mass
51
52 87 spectrometry (PTR-MS), gas chromatography ion mobility spectrometry (GC-IMS), atmospheric pressure
53
54 88 chemical ionisation- mass spectrometry (APCI-MS) and offline technologies including gas

1
2
3 89 chromatography mass spectroscopy (GC-MS) and comprehensive two-dimensional gas chromatography-
4
5 90 mass spectrometry (GCxGC-MS) will be utilised for VOC discovery and replication. For offline
6
7 91 technologies a standardised CE marked breath sampling device (ReCIVA[®]) will be used. All recruited
8
9 92 participants will be characterised using existing blood biomarkers including C - reactive protein (CRP),
10
11 93 brain derived natriuretic peptide (BNP), Troponin-I and blood eosinophil levels and further evaluated
12
13 94 using a range of standardised questionnaires, lung function testing, sputum cell counts and other
14
15 95 diagnostic tests pertinent to acute disease.
16
17

18 96 **Ethics and dissemination**

19
20
21 97 The National Research Ethics Service Committee East Midlands has approved the study protocol (REC
22
23 98 number: 16/LO/1747). IRAS 198921. Findings will be presented at academic conferences and published in
24
25 99 peer-reviewed scientific journals. Dissemination will be facilitated via a partnership with the East
26
27 100 Midlands Academic Health Sciences Network and via interaction with all UK funded MRC/EPSRC
28
29 101 molecular pathology nodes.
30
31

32 102

33
34
35 103 **Key words:** Breathlessness; Breath analysis; Volatile Organic Compound, Observational study
36
37

38 104

39
40
41 105

42
43
44 106

45
46
47 107

48
49
50 108

51
52
53 109

110

111

112

113

114

115 **Strengths and Limitations of this Study**

116

- A pragmatic real world, prospective, observational study across three admission units that focuses on the systematic discovery and replication of VOC in acutely breathless patients using both online and offline technologies

117

118

119

- The proposed study is the largest of its kind in acute disease to characterise VOC with a range of additional assessments that will build a comprehensive phenotype of acute cardio-respiratory exacerbations

120

121

122

- The proposed study will build an infrastructure for research and subsequent evaluation of VOC in interventional trials within acute cardio-respiratory exacerbations

123

124

- Prior acute treatment exposure will need to be accounted for when evaluating potential discriminative biomarkers

125

126

- VOC technologies are not currently suited for deployment in patients that are of high clinical acuity

127

128

-

129 **1. Introduction:**

130 Breathlessness is a common symptom of cardio-respiratory illnesses that has a significant direct impact

131 on patients' wellbeing as well as a substantial economic burden on healthcare systems [1]. Although its

132 etiologies can be variable, exacerbations of common complex chronic cardio-respiratory conditions

1
2
3 133 account for approximately 70% of acute presentations with breathlessness, namely exacerbations of
4
5 134 asthma and COPD, acute heart failure and community acquired pneumonia [2]. Moreover, moderate and
6
7 135 severe breathlessness is significantly associated with all-cause, cardiovascular and COPD mortality[3]. As
8
9 136 a consequence symptomatic breathlessness warrants rapid evaluation and targeted diagnostics at
10
11 137 presentation.
12

13
14 138
15
16 139 Diagnostic evaluation of acute breathlessness is heavily reliant on blood based biomarkers e.g. CRP,
17
18 140 BNP, Troponin and on occasions blood eosinophil levels. These biomarkers have clinical utility primarily
19
20 141 in patients with single pathologies, but have poor discriminatory power in patients with multifactorial
21
22 142 presentations of acute breathlessness[4]. There is therefore an unmet need for the development of
23
24 143 sensitive and specific biomarkers that differentiate acute breathlessness from its recovery and the
25
26 144 common cardio-respiratory conditions that present with acute breathlessness.
27

28
29 145
30
31 146 CRP plays an important role in diagnosing breathlessness caused by an underlying bacterial
32
33 147 pneumonia[5], as well as predicting mortality in patients with COPD [6]. BNP is routinely utilised in
34
35 148 acute settings to support the diagnosis of acute heart failure [7]. The European Society of Cardiology
36
37 149 (ESC) recommends BNP threshold values of <100 pg/mL to rule out acute congestive cardiac failure and
38
39 150 values > 500 pg/ml as diagnostic of acute exacerbations of heart failure [8].
40

41 151
42
43 152 The role of peripheral blood eosinophil count in airway inflammation was poorly understood up until the
44
45 153 second half of the 19th century when Paul Ehrlich, a German physician and Nobel prize winner,
46
47 154 introduced eosin in his technique for white cell differentiation in 1879[9]. Considerable advances in the
48
49 155 field of airway inflammation and the role of eosinophils have taken place since [10-12]. More recently
50
51 156 Bafadhel *et al* suggested that peripheral blood eosinophil count can be used to direct corticosteroid
52
53 157 therapy during COPD exacerbations in single centre study [13].
54
55
56
57
58
59
60

1
2
3 158 Currently, blood biomarkers together with clinical, physiological and imaging parameters are used in
4
5 159 diagnosing the cause of acute breathlessness. Blood biomarkers may be less specific as they originate far
6
7 160 from the target organs of interest (the heart and the lungs in cardio-respiratory disease). Sputum, although
8
9 161 potentially a more definitive lung specific matrix, is comparatively difficult to obtain particularly in
10
11 162 acutely unwell patients, limiting its use in acute disease and highlighting the need for better biomarkers.
12
13 163 Ideally these biomarkers would have the following characteristics, (i) they would originate from the target
14
15 164 organ of interest, (ii) they would significantly add value to conventional risk scoring and diagnostic
16
17 165 algorithms in acute breathlessness, (iii) they would be minimally invasive and suitable for rapid point of
18
19 166 care diagnosis in emergency rooms and acute admissions units (iv) they would have diagnostic value in
20
21 167 patients with multifactorial acute breathlessness.
22
23
24 168

25
26 169 Exhaled breath contains thousands of volatile organic compounds (VOC) that reflect biological processes
27
28 170 occurring in the host both locally in the airways and systematically offering the potential to develop more
29
30 171 effective biomarkers in acutely breathless patients (**Figure 1**).
31
32 172

33
34 173 The proposed program of research will use a combination of offline and online technologies to identify
35
36 174 and evaluate the diagnostic and prognostic value of VOC in patients with acute cardio-respiratory related
37
38 175 breathlessness (**Figure 2**).
39
40 176

41
42 177 Exhaled breath analysis of volatile chemicals offers the potential to develop more effective biomarkers in
43
44 178 acutely breathless patients. The use of breath analysis for disease diagnostics dates back to ancient Greeks
45
46 179 where physicians used exhaled breath to diagnose different diseases. Breath odours allow correct
47
48 180 associations to certain diseases. For example, the sweet smell of diabetic ketoacidosis, the fishy smell of
49
50 181 breath associated to liver illness, the urine-like odour of kidney disease and the smell of the breath of
51
52 182 patients with lung abscesses, caused by the proliferation of anaerobic bacteria [14-17].
53
54
55
56
57
58
59
60

1
2
3 183 More recently, exhaled breath analysis has demonstrated early proof of concept in the diagnosis of acute
4
5 184 heart failure, and ventilator associated pneumonia[18] . The validity of breath analysis has also been
6
7 185 demonstrated in breathless children[19]. This population is likely to prefer breath-based tests, as these are
8
9 186 minimally invasive. Importantly, a variety of point of care sensors are now available to evaluate potential
10
11 187 exhaled breath biomarkers in emergency care settings.
12

13
14 188

15
16 189 A study by Van Berkel et al demonstrated the ability to distinguish COPD subjects from controls solely
17
18 190 based on the presence of VOCs in breath, suggesting that analysis of VOC might be highly relevant for
19
20 191 diagnosis of COPD [20]. This established the basis of further studies of VOC in COPD [21-
21
22 192 25].recommending larger studies for validation.
23

24 193 Several other studies found that VOC profiling in diagnosing asthma is potentially feasible [26-32]. This
25
26 194 however has been done in relatively small numbers in stable disease.
27

28 195 Despite the novelty of non-invasive sampling technology and the growing interest in exhaled breath
29
30 196 analysis there remains a disappointing level of comparability across studies due to the lack of
31
32 197 standardisation and appropriate data analysis methods. A recent systemic review by Anders Christiansen
33
34 198 *et al* compared eleven publications reporting very heterogeneous designs, methods, patient group sizes,
35
36 199 data analytics and, consequently, quite varying results [33].
37
38

39 200

40
41 201 To our knowledge, no other large studies exploring the use of breath biomarkers in profiling acute
42
43 202 breathlessness have been completed. Several studies have explored the use of electronic nose (eNose) in
44
45 203 stable disease with good discriminatory power in COPD [34], Pneumonia [35] and heart failure[36] with
46
47 204 relatively small sample size. While eNose has now been widely used in detecting various VOC patterns,
48
49 205 GC-MS, a largely validated methodology, remains the gold standard technique for detecting VOCs in
50
51 206 exhaled breath. The focus of the current research study will be to evaluate acutely breathless cardio-
52
53 207 respiratory patients using a combination of ‘discovery’ and near-patient care breath sampling
54
55 208 technologies.
56
57
58
59
60

209 Medical Research Council (MRC) and Engineering and Physical Sciences Research Council (EPSRC)
210 have commissioned a series of molecular pathology nodes aimed at developing molecular signatures
211 relevant to disease diagnosis and progression. This was triggered by the clear need for alliance between
212 academic institutions, industry and NHS partners to enhance the benefits of stratified medicine for
213 patients[37, 38].

214 University of Leicester and Loughborough University were awarded a joint molecular pathology node
215 East Midlands Breathomics Pathology Node (EMBER) which this study forms a key part of.

217 **2. Methods and Analysis**

218 **2.1. Study design**

219 A prospective real world observational study across three acute admissions units within Leicestershire
220 (two adult admissions units and one children's assessment unit). The acute units routinely assess and treat
221 cardio-respiratory admissions due to breathlessness in adults and children.

222 Participants with self-reported acute breathlessness, either requiring admission or a change in baseline
223 treatment, will be screened for the study. Informed consent will be obtained in all participants following a
224 clinical review by a senior decision maker within 24 hours of acute admission (**Figure 3**).

227 **2.2. Objectives**

228 **2.2.1. Primary objective**

- 230 • To evaluate the sensitivity, specificity, positive and negative predictive value of exhaled breath
231 VOC biomarkers to differentiate acute breathlessness in cardio-respiratory patients.

232 **2.2.2. Secondary objectives:**

- 234 • To replicate selected breath VOC biomarkers identified in acute breathlessness.

- 1
2
3 235 • To discover and replicate breath VOC biomarkers that differentiate the common cardio- respiratory
4
5 236 conditions that cause acute breathlessness, specifically (i) acute heart failure, (ii) community
6
7 237 acquired pneumonia, (iii) adult exacerbations of asthma and COPD and age-matched adults that do
8
9 238 not have cardio-respiratory disease or breathlessness.
10
11 239 • To quantify the level of clinical uncertainty in the primary diagnosis using a 100 mm visual
12
13 240 analogue scale and independent clinical adjudication of case notes blinded to the following blood
14
15 241 biomarkers (i) CRP (ii) BNP (iii) Troponin-I (iv) blood eosinophils, but not clinical history and
16
17 242 acute presentation nor chest x-ray imaging. Potential discriminatory breath VOC biomarkers will
18
19 243 be adjusted for clinical uncertainty in statistical models.
20
21 244 • To identify and replicate exhaled breath VOC biomarkers in school age children treated in hospital
22
23 245 for severe asthma attacks and compare these to age-matched healthy controls.
24
25

26 246 **2.2.3. Exploratory end points (where applicable):**
27 247

- 28
29 248 • To evaluate the dynamic profile of selected breath VOC between the acute state and the recovery
30
31 249 state post exacerbation.
32
33 250 • To evaluate the relationship between exhaled VOC biomarkers and clinical outcomes including
34
35 251 (i) hospital readmission at 30 and 60 days post event (ii) all-cause mortality over a 2 year period
36
37 252 post admission.
38
39 253 • To evaluate the relationship between breath VOC biomarkers and functional measures
40
41 254 e.g. physical performance and activity
42
43 255 • To explore potential breath VOC biomarkers of multifactorial acute breathlessness
44
45 256 • To evaluate the relationship between diet, lifestyle and environment upon breath VOC
46
47 257 biomarkers
48
49
50

51 258
52 259 **2.3. Sample size estimation**
53
54 260
55
56
57
58
59
60

261 Preliminary data was utilised to conduct sample size estimates from a cohort of acutely breathless patients
262 admitted to acute admissions units over a 6 month period (February 2017 to August 2017). Hundred and
263 twelve adult participants (asthma 46, community acquired pneumonia 26, COPD 22) and eighteen healthy
264 controls were utilised for the analysis.

265 A panel of ten pre-specified aldehydes, based on literature search [31, 39, 40], were extracted from breath
266 using gas chromatography-mass spectrometry (GC-MS). The aldehydes were normalised to a common
267 internal standard and were not background-subtracted.

268 A closed formula from Hsieh *et al*[41], relating sample size to observable effect size, was used to
269 calculate sample size from logistic regression models of the ten aldehydes with acute breathlessness as the
270 outcome measure. The sample size estimates are also relevant to acute class comparisons vs. the sum of
271 other acute classes.

272 Based upon the sample size estimates we would have an 80% power, with a type 1 error rate of 5%, to
273 detect an odds ratio of association of 1.2 between two disease classes with 55 patients per class. Given the
274 fact that study seeks to discover and replicate breath VOC amongst five adult disease classes (community
275 acquired pneumonia, heart failure, COPD, asthma and healthy aged matched subjects) we would require
276 110 adult patients per class – 550 patients across the program to achieve these aims.

277 The closed formulae by Tihaki *et al*,[42] were also utilised to understand the discriminatory power that
278 the samples sizes above would provide with respect to biomarker sensitivity and specificity; The
279 following assumptions were made:

- 280 • That a sensitivity of 80 % with a precision of 5% would provide a useful biomarker capable of
281 'ruling out' an acute class. The same target was applied to specificity.
- 282 • We assume a prevalence of acute breathlessness of 80% as the recruitment campaign uses acute
283 breathlessness as the initial stratification tool for recruitment and 1:5 patients recruited will be
284 non-breathless healthy controls

285 • We aim to balance group sizes across classes equally

286 For a type 1 error rate of 0.05 and a 95% confidence interval

287 $N_{\text{sensitivity}} = 307$

288 $N_{\text{specificity}} = 1,230$

289

290 For a type 1 error rate of 0.05 and a 90% confidence interval

291 $N_{\text{sensitivity}} = 218$

292 $N_{\text{specificity}} = 871$

293

294 For a type 1 error rate of 0.05 and an 85% confidence interval

295 $N_{\text{sensitivity}} = 166$

296 $N_{\text{specificity}} = 664$

297

298 For a type 1 error rate of 0.05 and an 80% confidence interval

299 $N_{\text{sensitivity}} = 131$

300 $N_{\text{specificity}} = 524$

301

302 Therefore, we are powered to identify sensitive biomarkers ($\geq 80\%$) of acute breathlessness with a
303 maximum marginal error in the estimate for sensitivity not exceeding 5% with 95% confidence. Similarly,
304 we are powered to identify specific biomarkers ($\geq 80\%$) of acute breathlessness with a maximum
305 marginal error in the estimate for specificity not exceeding 5% with 80% confidence.

306 For the primary analysis the outcome will be treated as a nominal variable with levels (i) acute heart failure
307 (ii) community acquired pneumonia (iii) adult exacerbations of asthma and COPD and (iv) acute
308 exacerbations in school age children treated in hospital for severe asthma attacks.

309 The relationship between the primary outcome and the exhaled breath VOC biomarkers will be modelled
310 using multinomial logistic regression. In addition to metabolomics markers the following independent
311 variables will be included in the model: clinical uncertainty score on a 100 mm VAS scale, age, and a

1
2
3 312 validated co morbidity score (the Charlson comorbidity score)[43, 44].
4

5 313
6
7 314 Receiver operator analyses will be utilised to generate ROC curves for individual and multiple panels VOC
8
9 315 predictors in the primary analysis.
10

11 316
12
13
14 317 To understand the dynamic profile of breath biomarkers during (i) the acute state and (ii) in the chronic
15
16 318 state up to 6 months post exacerbation, a repeated measures model with a random intercept and random
17
18 319 effect for time will be fitted, the random effects will be fitted for each patient. For the repeated measures
19
20 320 mixed model an unstructured covariance will be assumed. To evaluate the relationship between breath
21
22 321 biomarkers and hospital readmission at 30 and 60 days Cox proportional hazards and frailty models will
23
24 322 be utilised [45]. Analysis of Multivariate Survival Data, [CITE] competing risk models and joint models
25
26 323 will be fitted [46]. Relationship between death and breath biomarkers will be evaluated using a logistic
27
28 324 regression model. Changes in outcome measures will be measured appropriately for each variable (e.g.
29
30 325 paired t-test, Mann-Whitney, repeated measures analysis). Tables of descriptive statistics will be compiled
31
32 326 for all key variables
33
34
35

36 327 All analysis will be performed using R 3.5.0 <https://www.r-project.org/>.
37

38 328 39 329 **2.4. Discovery and Replication studies**

40
41
42 330 Specific indicator conditions have been selected for targeted recruitment according to their high
43
44 331 prevalence and unmet need, their high morbidity and mortality and the need to develop better diagnostic
45
46 332 and prognostic algorithms in acute care pathways.
47

48
49 333 The indicator diagnoses of interest are (i) exacerbations of adult asthma and COPD, (ii) community
50
51 334 acquired pneumonia (iii) acute heart Failure and (iv) exacerbation in school age children treated in hospital
52
53 335 for severe asthma attacks.
54
55
56
57
58
59

336 Patient level clinico-pathological and outcome data (spanning the entire acute pathway) will be collected
337 in parallel to breath sampling. In addition, breath samples will be acquired in the stable state post
338 exacerbation (**Figure 3**).

339
340 Age matched healthy volunteers will be recruited where possible at separate visits. For the purposes of this
341 study, healthy volunteers will be defined as participants who have no prior history of asthma, COPD, heart
342 failure and have not been admitted to hospital with community acquired pneumonia within 6 weeks of the
343 baseline study visit. For acute admission the study team will approach the spouse, parent or sibling of the
344 index case and seek informed consent for study assessments. All healthy subjects will undergo two
345 assessments separated by a duration of 8-16 weeks to match the acute and recovery time points elapsed in
346 their index case/partner/spouse/sibling/child. Additional healthy volunteers will be identified from local
347 recruitment databases and via advertising

349 **2.4.1. Discovery Phase (Project months 1-24):**

350 The aim of the discovery phase is to discover putative discriminatory breath VOC, using both offline and
351 online technologies.

352 Pre-planned recruitment of acutely breathless patients will be enriched into the following disease strata
353 following senior clinical decision maker assessment and within 24 hours of acute admission.

354 Acute adult heart failure (n=55), adult community acquired pneumonia (n=55), adult exacerbations of
355 asthma (n=55) and COPD (n=55), acute severe asthma attacks in school age children (n=50).

356 Additional age matched healthy volunteers (n=55 adults and 50 children) will be identified as a non-
357 disease reference group (**Table 1**).

358

359

360 **2.4.2. Replication Phase (years 3-4)**

361 The aim of the replication phase is to replicate putative discriminatory breath VOC /VOC signatures
 362 identified in the discovery phase.

363 Similar to the discovery phase, recruitment of acutely breathless patients will be enriched into the
 364 following disease strata following senior clinical decision maker assessment and within 24 hours of acute
 365 admission.

366 Acute adult heart failure (n=55), adult community acquired pneumonia (n=55), adult exacerbations of
 367 asthma (n=55) and COPD (n=55), acute severe asthma attacks in school age children (n=25) (**Table 1**).

368 Additional age matched healthy volunteers (n=55 adults and 25 children) will be identified as a non-
 369 disease reference group.

370 **Total combined sample size of the discovery and replication phases = 700 participants**

Disease Category	Discovery	Replication
Acute Adult Asthma	55	55
Acute COPD	55	55
Acute Heart Failure	55	55
Community Acquired Pneumonia	55	55
Adult healthy volunteers	55	55
Acute paediatrics Asthma	50	25
Paediatrics healthy volunteers	50	25
Total sample	375	325

371 **Table (1): Table summarising recruitment targets for both adult and paediatric groups.**

372 **2.5. Schedule of assessments**

373 A schedule of acute assessments is outlined below and aligns to the movement of acute patients through
 374 the clinical care pathway and the overall aim of developing a complete phenotypic picture of acutely
 375 breathless patients.

377 **2.5.1. Defining acute breathlessness**

378 At presentation (within 24 hours of admission) to one of three acute admissions units potentially eligible
 379 patients will be identified following confirmation of acute breathlessness, identified as (i) patient defined
 380 acute breathlessness and/or (ii) 1 unit increase above patient reported baseline in the extended medical

381 research council (eMRC) dyspnoea score [47, 48] and at least one of the indicator diagnoses identified as
 382 the primary clinical diagnosis by a senior clinical decision maker.

383 **2.5.2 Informed consent**

384 Patients meeting the pre-specified definition of acute breathlessness will be approached for informed
 385 consent in to the breath VOC biomarker study. Only patients that are eligible to give full written informed
 386 consent will be recruited.

387 **2.5.3 Collection of blood based pathology markers**

388 Collection of the blood biomarkers CRP, BNP, Troponin-I and blood eosinophil count will be performed
 389 both acutely and following recovery, when not taken as part of clinical care pathway. These are currently
 390 used in profiling acutely breathless patients in clinical practice (**Table2**).

391

Test	ANALYSER/METHOD	LOWER LIMIT OF DETECTION	UPPER LIMIT OF DETECTION
C-Reactive protein (CRP)	Siemens Advia Chemistry XPT, PEG enhanced immunoturbidimetric. Siemens Advia 1800, PEG enhanced immunoturbidimetric	5 mg/L	Diluted to result
B-type natriuretic peptide (BNP)	Siemens Advia Centaur XPT, two-site sandwich immunoassay using direct chemiluminescent technology	2.0 pg/mL	1445 pg/mL
Troponin-I	Abbott Architect i2000SR, three-site sandwich immunoassay using direct chemiluminescent technology (CMIA).	5.0 ng/L	50,000 ng/L

392 **Table (2):** Type of analyser and methodology used for blood biomarker calculation. The table outlines
 393 analyser make, methodology, upper and lower limits of detection as per the University Hospitals of
 394 Leicester NHS Foundation trust laboratory guidelines.

395

396 **2.5.4 Breath VOC sampling**

397 Offline breath sampling using GC-MS coupled with a standardised and CE marked breath sampler-
 398 ReCIVA[®][49] and comprehensive two-dimensional gas chromatography-mass spectrometry, coupled
 399 with a standardised and CE marked breath sampler (ReCIVA[®] GCxGC-MS) will be performed. Gas
 400 chromatography is considered a gold standard technique in detecting volatile organic compounds and as

401 such its sampling will be prioritised. Additionally the following online technologies, proton transfer
 402 mass spectroscopy (PTR-MS), gas chromatography - ion mobility spectroscopy (GC-IMS) and
 403 atmospheric pressure chemical ionisation- mass spectrometry (APCI-MS) will be evaluated according to
 404 the sampling strategy outlined in section 3, **Figure 3 and Table 3.**

	COPD		Asthma		Pneumonia		Heart Failure		Healthy		Paediatrics	
Time point	1	2	1	2	1	2	1	2	1	2	1	2
Written informed consent	x		x		x		x		x		x	
Volatile organic compound (VOC) sampling												
ReCIVA- gas chromatography and mass spectrometry (GC-MS)	x	x	x	x	x	x	x	x	x	x	x	x
ReCIVA- comprehensive two-dimensional gas chromatography (GCxGC-MS)	x	x	x	x	x	x	x	x	x	x	x	x
Atmospheric pressure chemical ionisation- mass spectrometry (APCI-MS)	x	x	x	x	x	x	x	x	x	x	x	x
Proton transfer reaction mass spectrometry (PTRMS)		x		x		x		x	x	x	x	x
Gas chromatography - ion mobility spectrometry (GC-IMS)	x		x		x		x				x	x
Pathology blood tests												
Full blood count (including differential cell count)	x	x	x	x	x	x	x	x	x	x	x	x
Brain natriuretic peptide (BNP) [pg/mL]	x	x	x		x		x	x	x			
Troponin-I [ng/L]	x		x		x		x		x			
C-Reactive protein (CRP) [mg/L]	x	x	x	x	x	x	x	x	x	x	x	x
Lung function tests											x	x
Hand held forced oscillation technique (FOT)	x	x	x	x	x	x	x	x	x	x		
Fractional exhaled nitric oxide (FeNO) - Flow rate 50 [ml/s]			x	x							x	x
Spontaneous sputum sample	x	x	x	x	x	x	x	x	x	x		
Bio-banking (urine, serum, plasma. sputum supernatants & plugs)	x	x	x	x	x	x	x	x	x	x	x	x

Transthoracic echocardiography	x						x					
--------------------------------	---	--	--	--	--	--	---	--	--	--	--	--

405 **Table (3):** Summary of baseline and follow up assessments. The table summarises key assessments
 406 carried out at different time points during the study. The participants may undertake any combination of
 407 the investigations listed at any of these time points.

408

409

410 **2.5.5. Collection of additional samples for future biomarker campaigns**

411 Collection of additional biomarkers for future biomarker discovery campaigns including (i) a urine
 412 sample, (ii) blood samples, up to 85mls for DNA, RNA, plasma and serum and peripheral blood cell flow
 413 cytometry in selected subjects and (iii) spontaneous or induced sputum samples (plugs and supernatants)
 414 will be carried out (**Table 3**).

415

416 All samples will be collected at time point 1 and at 2 (**Figure 3**). The additional samples will be used for
 417 future omics analyses, these may include detailed analysis of the metagenome in sputum and proteomics
 418 applied to urine and serum samples.

419

420 **2.5.6. Physiological characterisation**

421 Physiological measures of lung function will be performed in acutely ill participants and at
 422 recovery including, (i) Hand held forced oscillation technique (FOT): an easily accessible
 423 measure of lung function. Patients favour this to spirometry as it is effort independent, unlike
 424 spirometry and requires less than a minute of quiet tidal breathing to obtain triplicate high quality
 425 measurements [50], (ii) Fractional exhaled nitric oxide (FeNO): A measure of airway
 426 inflammation in asthmatic patients[51, 52] (iii) Echocardiography: Two dimensional
 427 transthoracic echocardiography will be performed in heart failure and COPD patients using an iE
 428 33 system with S5-1 transducer (frequency transmitted 1.7 MHz, received 3.4 MHz; Philips
 429 Medical Systems, Best, The Netherlands). Standard techniques as per American Society of

1
2
3 430 Echocardiography guidelines (ASE)[53] were used to acquire 2D, colour and Doppler images in
4
5 431 conventional parasternal long-axis, short-axis and apical 4-, 2- and 3-chamber views. Left
6
7 432 ventricular ejection fraction (LVEF) was calculated using the biplane method of discs formula
8
9 433 (Simpson's rule) to derive left ventricular volume indices.
10
11
12

13 434 All participants are encouraged to report any testing related discomfort or concerns to the research team to
14
15 435 terminate the sampling process.
16

17 436 **2.5.7 Recovery follow up**

- 19 437 • Patient recovery will be defined as:
20 438 (i) Patient reported recovery from the acute exacerbation spell and back to their
21 439 baseline extended MRC score or clinician defined recovery from the acute
22 440 exacerbation spell
23 441 and
24 442 (ii) At least 6 weeks post exacerbation event (up to 6 months).
25 443

26 444 Patients that re admit to hospital between visits 1 and 2, can have additional visit 1 assessments. Visit 2
27 445 will be taken as recovery following the subsequent admission. If a patient is admitted to hospital after visit
28 446 2 then they will be eligible to be recruited as a new study participant.
29 447

30 448 The schedule of assessments at the recovery visit is outlined (**Table 3**).
31
32

33 449 **2.6. Clinical Adjudication:**

34 450
35 451 In an effort to reduce data variability and minimise bias, an independent panel consisting of two senior
36 452 acute clinicians (SS & NG) will review all pertinent clinical and diagnostic source documentation, whilst
37 453 blinded to admission blood biomarkers and clinical diagnosis.
38 454

39 455 All acutely breathless adult patients' notes will be sequentially divided for adjudication. The panel will
40 456 independently determine the primary diagnosis of highest probability from a list of the four potential acute
41 457 indicator diagnoses in section 2.4 and mark their level of clinical certainty on a 100 mm visual analogue
42 458 scale (VAS scale). The panel members will be able to review imaging, electrocardiograms (ECGs), and
43 459 other relevant information but not admission blood based pathology tests.
44 460

45 461 In a subset of patients adjudication will be validated by separate panel member to ensure between observer
46 462 agreement using Bland-Altman analysis and inter-rater agreement of the primary diagnosis using Kohen's
47 463
48 464
49 465
50 466
51 467
52 468
53 469
54 470
55 471
56 472
57 473
58 474
59 475
60 476

462 kappa via repeated evaluation of a subset of cases (see statistical methods section 2.3).

463

464 **2.7. Clinical Informatics:**

465

466 Clinical data collection will be undertaken using a securely-hosted bespoke database system (ADD)
467 developed within the NIHR Leicester Biomedical Research Centre – Respiratory (BRC). The system links
468 acute admission episodes to hospital pathology records; historical respiratory physiology tests; and
469 demographic information. The system provides functionality to validate data entry; manually verify
470 records and highlight incomplete records. A custom VOC ‘module’ has been created to support data
471 collection within the study visits (1 and 2), and standardise diagnoses and medications through the use of
472 clinical ontologies as well as linking hospital records/tests to patient visits.

473

474 Non-ADD based clinical data (e.g. hospital admissions, re-admissions and mortality) will be extracted
475 from the hospital data warehouse using identifiable patient identifiers, and subsequently pseudonymised
476 prior to integration.

477

478 An informatics pipeline will be created to facilitate the transfer of chemo-metric data from remote
479 computers to the data repository. This will include tools to (i) enforce the correct labelling of data sets
480 (e.g. study number, visit, type/source of sample) prior to automated validated transfer to the repository;
481 (ii) record information about the sample process; (iii) search and extract data sets from the repository for
482 subsequent analysis. Prior to analysis, clinical and chemo-metric data will be integrated using the study
483 number and any potentially identifiable information will be removed.

484

485 **3. Breath profiling**

486

487 The technologies utilised in the VOC study during discovery and replication phases are:

1
2
3 488 Offline technologies

4 489 - ReCIVA+ GC-MS

5 490 - ReCIVA + GC x GC-MS

6 491 -

7
8
9
10 492 Online technologies

11 493 - GC-IMS

12 494 - PTR-MS

13 495 - APCI-MS

14
15
16 496

17
18
19
20 497 Offline technologies will underpin the discovery analyses owing to their ability to identify chemical
21
22 498 identity and their recognition as the analytical gold standard in exhaled breath VOC analysis [54].

23
24 499 In contrast online technologies will be utilised for VOC biomarker replication and at the recovery visits
25
26 500 owing to their portability and potential for future point of care testing. **(Figure 4)**.

27
28 501 A brief description of the core VOC platforms is provided below

29
30
31 502 A CE marked breath sampling device (ReCIVA®) developed by Owlstone Medical, will be used to
32
33 503 sample breath onto two adsorbent Tenax tubes. Participants will be asked to breathe through the ReCIVA
34
35 504 face mask for a maximum of 300 seconds, aiming for collection of $\geq 80\%$ of the target sample volume of
36
37 505 1 litre, after which the Tenax tubes will be transferred to the laboratory for analysis. This effectively
38
39 506 allows de-coupling of the breath sampling from the breath sensor and analysis platforms in selected
40
41 507 patients that are not able to mobilise to a real time breath sampling device. The Owlstone ReCIVA
42
43 508 sampler will be utilised in breath collection for offline technologies namely GC-MS and GCxGC-MS.
44
45 509 The ReCIVA sampler is capable of entraining oxygen and is therefore suitable for patients with mild
46
47 510 respiratory failure requiring low flow rates of oxygen to maintain target oxygen saturations [49].

48
49
50
51 511 **3.1. Gas chromatography and mass spectroscopy (GC-MS):** is a commonly applied methodology used to
52
53 512 accurately measure trace gases in complex mixtures such as exhaled air [54]. Pre-concentrating breath
54
55 513 volatiles by various means and subsequent analysis constitute a reliable and sensitive method for VOC

1
2
3 514 analysis [55]. Despite its high sensitivity, it is however, a time consuming technique and carries a risk of
4
5 515 contamination at the pre-concentration step. It is also not suitable for online and multiple measurements
6
7 516 limiting its use as a point-of-care testing technology for VOC [56].
8
9

10 517 **3.2. Comprehensive two-dimensional gas chromatography-mass spectrometry (GCxGC-MS):** an
11
12 518 advanced analytical technique for the analysis of complex organic matrices; its main advantage is the
13
14 519 unparalleled separation power it affords over conventional one-dimensional chromatographic techniques
15
16 520 [57]. Previous research, albeit sparse, has demonstrated the potential of GCxGC-MS for breath analysis
17
18 521 with the number of VOC detected exceeding those detected by conventional GC-MS [58, 59]. GCxGC-
19
20 522 MS of breath metabolites has been used for the identification of biomarkers related to glucose metabolism
21
22 523 [60, 61], tuberculosis [62] and radiation response [63]. This has generated interest within the breath
23
24 524 research community, however, such studies were conducted on a small scale (<50 patients) and involved
25
26 525 the use of expensive detectors and modulators. Method development and analysis of the data-rich GCxGC
27
28 526 chromatograms, however, can be time-consuming and require specialist knowledge.
29
30

31
32 527 **3.3. Proton-transfer-reaction mass spectrometry (PTRMS):** a real time technique, capable of
33
34 528 simultaneously measuring the evolution of multiple gas metabolites from a single breath. It has been used
35
36 529 for the identification of potential useful VOC biomarkers for diagnosis of a variety of diseases: including
37
38 530 various cancers[64-66], liver disease[67, 68] and respiratory disease[69]. It has several advantages in
39
40 531 clinical settings, such as the speed of sampling, the instant result achieved and the lack of need for sample
41
42 532 storage or shipping. However, owing to the lack of pre-concentration or chromatographic separation,
43
44 533 sensitivity and definitive compound identification can be somewhat limited when compared to GC-MS.
45
46
47 534 Two breath sampling devices will be used. The first device is a Loccioni SOFIA GSI-S; the subject is
48
49 535 required to exhale a single breath, five times (three if providing five samples proves too difficult) into a
50
51 536 sterile mouthpiece connected to an electrostatic bacterial/viral filter whilst wearing a nose clip (all CE
52
53 537 marked). Flow from the mouthpiece passes into a gas sampling interface capnograph (Loccioni GSI-S –
54
55 538 CE marked) and real-time user feedback of flow is provided on screen, allowing the regulation of the
56
57
58
59
60

1
2
3 539 breath sampling rate. The gas sampling interface acts to simultaneously trigger the acquisition of the
4
5 540 PTR-ToF-MS data and the exhaled breath travels through the capnograph down a heated sample line into
6
7 541 the ion source of the PTR-ToF-MS
8
9

10 542 The second breath sampling device is a ReCIVA breath sampler (Owlstone) with one of the adsorbent
11
12 543 Tenax tubes replaced with an outlet tube adapted for online sampling. The exhaled breath is transferred to
13
14 544 the PTR-ToF-MS via a heated transfer line connected to the outlet tube, continuously drawn at a constant
15
16 545 flow rate by the PTR-ToF-MS. The online adaptation of the consumable adsorbent tube does not affect
17
18 546 the CE mark of the ReCIVA sampling device.
19
20

21 547 Once the breath sample reaches the PTR-ToF-MS, via either breath sampler, the breath mixes with
22
23 548 protonated water (H_3O^+) inducing proton transfer to the target volatile organic compounds (VOCs)
24
25 549 present, resulting in their ionisation. Sample ions are then guided into the time of flight mass
26
27 550 spectrometer and mass spectra, showing the abundance and mass of the VOCs present, are collected
28
29 551 throughout the exhalation. Following sampling, mouthpieces, filters and nose clips are disposed of and all
30
31 552 patient contacted surfaces wiped down with antiseptic cleaning wipes in preparation for the next patient.
32
33
34
35

36
37 553
38 554 **3.4. Gas chromatography- ion mobility spectrometry (GC-IMS) (B&S Analytik):** Allows the detection
39
40 555 of volatile organic compounds down to ultra-trace level (ng/L - to pg/L - range). For years, IMS has been
41
42 556 used to discover potential discriminatory breath VOC in lung cancer [70, 71], COPD[72, 73] and
43
44 557 asthma[73]. Sampling takes place using a Spiroscout spirometer. The patients exhale through a disposable
45
46 558 mouth piece connected to a Teflon tube. A piezoelectric pressure sensor is used to monitor the breathing
47
48 559 profile, this opens the sampling valve at the appropriate point in the breath profile to collect end-tidal
49
50 560 breath in a sample loop of 10 mL volume. After filling this loop, the collected sample air is then
51
52 561 transferred to a multicapillary column for a chromatographic separation, which is achieved in 12 min. The
53
54
55
56
57
58
59

1
2
3 562 separated molecules are then transferred into the IMS, ionised and then separated according to their
4
5 563 mobility in a weak electric field.
6
7 564 The technology's multiple advantages of ultra-sensitivity, portability, online sampling and short analysis
8
9 565 time (typical analysis time of 10 minutes) with real time detection, brings a promise to provide immediate
10
11 566 and potentially reliable results for point of care breath diagnostics. Another concept with IMS devices is
12
13 567 that once the required breath signatures have been discovered using GC-MS, IMS offers the potential to
14
15 568 be 'tuned' for selective detection of VOC.

16
17
18 569 **3.5. Atmospheric pressure chemical ionisation- mass spectrometry (APCI-MS) semiportable compact**
19
20 570 **version (Advion):** is one of less sensitive but more affordable versions of mass spectrometers released to
21
22 571 the commercial market in recent years. The device uses APCI to produce ions. Although the most
23
24 572 common use of APCI-MS systems is the detection in liquid chromatography applications, the technique
25
26 573 has proven to be a valuable tool for direct measurement of VOC in air[74, 75] food[76, 77] and breath[78,
27
28 574 79]. Recently, the technique has shown potential for online, real time profiling of pseudo-metabolites in
29
30 575 exhaled breath [80] with sensitivity comparable with other techniques. By combining miniaturised MS
31
32 576 technology with APCI techniques, adequate quality of on-site, real time measurements with minimal or
33
34 577 no sample preparation requirement can be provided. This is a desirable outcome as it overcomes main
35
36 578 limitation of using standard breath analysis method in clinical setting, which is a need for breath sample

37
38 579 ~~There remains an overall lack of standardisation and rigour across these technologies which hindered~~
39
40 580
41
42 581
43 608 There remains an overall lack of standardisation and rigour across these technologies which hindered
44
45 609 previous advancements in breath discovery; something we intend to minimize.

47 610 **4. Chemometric processing and data analysis:**

48
49
50 611 GC-MS breath data will be aligned, deconvoluted and the features for each participant will be extracted.
51
52 612 The extracted features will be grouped and classified by retention index and mass spectrum. The
53
54
55
56
57
58
59
60

1
2
3 613 registered and aligned data will be linked to participant meta-data to generate a breath matrix. Data
4
5 614 handling and analysis will be performed by a senior statistician.
6
7

8 615 The breath matrix is a $n \times p$ matrix where n is the number of subjects and p is the number of VOC. The
9
10 616 breath matrix is high dimensional with $p \gg n$ and many potentially correlated VOC. In view of this, we
11
12 617 will employ sparse partial least squares discriminant analysis (sPLS-DA)[81] to investigate which of the
13
14 618 VOC can identify breathlessness. We will also investigate which of the VOC can discriminate between
15
16 619 the different disease states including acute exacerbations of asthma and COPD and Pneumonia. In
17
18 620 addition to the supervised methods, unsupervised methods will be explored, specifically sparse principle
19
20 621 component analysis (sPCA)[82].
21
22

23
24 622 Extracted VOC will also be investigated. Relationships between VOC and patient reported acute
25
26 623 breathlessness will be analysed using logistic regression model. VOC associated with patient associated
27
28 624 acute breathlessness will be incorporated into multinomial logistic regression models in conjunction with
29
30 625 CRP, BNP, blood eosinophils and Troponin-I, pathology biomarkers currently in use for diagnosing
31
32 626 undifferentiated breathlessness. In addition to the conventional binary and multinomial logistic regression
33
34 627 models, regression models [83].
35
36
37

38 628 **5. Ethics and dissemination:**

39
40
41 629 The study has obtained full ethical approval from the London South East Research ethics Committee, REC
42
43 630 reference 16/LO/1747. IRAS project ID 198921. Publications will be prepared according to the MRC-
44
45 631 EMBER consortium agreement and the University of Leicester publications policy. All intended
46
47 632 publications will be submitted to the EMBER executive board for review and comments within 60 days of
48
49 633 journal submission. Authorship will be according to contribution and internationally recognised guidance
50
51 634 on journal authorship.
52
53

54
55 635 **6. Study dates:** 01/2/2017 – 30/10/2020
56
57
58
59
60

636 **7. Authors' contributions:**

637
638 S.S, C.E.B, N.Gr, P.Th and P.Mo conceived the study, obtained funding, wrote the study protocol,
639 obtained ethical and MHRA approvals for the study and coordinated the deployment of analytical testing
640 methods for breath analysis. W.I took the lead in writing the manuscript with support from S.S. Planning
641 and recruitment of adult participants was carried out by W.I, S.Jo, B.Pa, A.Aw, R.Ph, G.Fo, A.Yo, R. J. R
642 and C.Wh. Paediatrics study design was conceived by E.Ga and C.Be and participants recruited by T. Mc
643 and C. Fo. Analytical chemistry team formed of M.Wi, R.Co, D.Sa, D.Ru and L.Br expertly handled all
644 the breath samples and planned an analysis structure. M.Ri, a senior statistician, constructed a statistics
645 and data analysis plan in conjunction with SS. Bioinformatics pipeline and electronic CRFs developed by
646 R.Fr and B.Zh. All authors, including R.Pe, H. Bh, B.Ha, A. Si, K. Ry, H. Pa, T. Su, L. L Ng, contributed
647 to the study design and study protocol.

648 **8. Protocol version:** Version 4, 1st April 2018

650 **9. Public and patient involvement:**

651 A series of consultations have taken place with our patient involvement team within the NIHR Biomedical
652 Research Centre (Respiratory Theme) and across the wider BRC PPI group. Representations from the
653 paediatrics team were also present. This group was sent copies of the participant documentation for review and
654 discussion. Various revisions have been made following on from these discussions.

655 **10. Funding statement:**

656 Research funding was provided by the MRC and EPSRC Stratified Medicine Grant for Molecular
657 Pathology Nodes (Grant No. MR/N005880/1). The research was supported by the National Institute for
658 Health Research (NIHR) Leicester Biomedical Research Centre and NIHR, Leicester Clinical Research

1
2
3 659 Facility and the Midlands Asthma and Allergy Research Association (MAARA) to whom we are
4
5 660 extremely grateful. The authors would like to acknowledge the invaluable efforts of the research nurses
6
7 661 responsible for the in-clinic sample collection as well as the input from the wider EMBER consortium
8
9 662 (Members list can be found at: <https://ember.le.ac.uk/web>). The views expressed are those of the
10
11 663 author(s) and not necessarily those of the NHS and NIHR or the Department of Health
12
13
14
15 664

16
17 665 **11. Competing interests:** SS has performed advisory services for Owlstone Medical.
18
19 666
20
21 667
22
23 668
24
25 669
26
27 670
28
29 671
30
31 672
32
33 673
34
35 674
36
37 675
38
39 676
40
41 677
42

41 678 **Figure legends:**

43 679 **Figure (1)** Relationship between lung proximity and degree of invasiveness of different lung matrices.

45 680 The figure plots the level of invasiveness of various lung matrices in relation to their proximity to the
46
47 681 lung. Given their pathological relevance, the degree of invasiveness of bronchoalveolar lavage (BAL) and
48
49 682 lung biopsy makes them less favourable in diagnosing respiratory diseases.
50
51
52
53
54
55
56
57
58
59
60

1
2
3 683 **Figure (2)** Multi-instrument use in breath sampling. Figure illustrates the various combinations of offline
4
5 684 and online devices used in breath sampling and the relevant pros and cons. Offline and online
6
7 685 technologies are used for the discovery and validation phases of the study respectively.
8
9

10 686 **Figure (3)** Study flow chart. Figure outlines the patient journey from admission through to discharge and
11
12 687 follow up. Participants with self-reported acute breathlessness presenting to University Hospitals of
13
14 688 Leicester are recruited within 24 hours following a senior decision maker review. Breath sampling is
15
16 689 carried out on the first visit, at recruitment, and the second visit, up to 6 months after discharge. Patients
17
18 690 are admitted through the standard operational emergency medical streaming and care pathways at the
19
20 691 University Hospitals of Leicester NHS Trust. Early outcomes (hospital readmission) are measured at 30
21
22 692 and 60 days and late outcomes (mortality) including respiratory and all-cause mortality are measured at 2
23
24 693 years. Assessments carried out at each time point are summarised in **Table 1**.
25
26
27

28 694 **Figure (4)** Multi-instrument use in breath sampling. Operational space of the analytical technologies
29
30 695 used in EMBER for the analysis of volatile organic compounds (VOC) in exhaled breath, including
31
32 696 proton transfer reaction-mass spectrometry (PTR-MS), atmospheric pressure chemical ionisation-mass
33
34 697 spectrometry (CMS), gas chromatography-ion mobility spectrometry (GC-IMS), gas chromatography-
35
36 698 mass spectrometry (GC-MS) and two-dimensional gas chromatography-mass spectrometry (GCxGC).
37
38 699 Comparing the typical molar mass range detectable; selectivity in detection owing to the type of
39
40 700 ionisation involved and the proton affinity of analytes; and the inclusion of a chromatographic separation
41
42 701 affecting total time of analysis. The online technologies involving chemical ionisation (PTR-MS, CMS
43
44 702 and GC-IMS) can be used in-clinic owing to short analysis times, but only detect lower molar mass
45
46 703 molecules with a proton affinity higher than 697 KJ mol⁻¹. Offline chromatographic techniques (GC-MS
47
48 704 and GCxGC) detect a wider range of compounds independent of proton affinity; however, the techniques
49
50 705 have longer analysis times and involve sample transportation and storage.
51
52
53

54 706
55
56
57
58
59
60

1
2
3 707
4
5 708
6
7 709
8
9 710
10
11 711
12
13 712
14
15 713
16
17 714
18
19 715
20
21 716
22
23 717
24
25 718
26
27 719
28
29 720
30
31 721
32
33 722
34
35 723
36
37 724
38
39 725
40
41 726
42
43 727
44
45 728
46
47
48
49
50
51 729
52 730
53 731
54 732
55
56
57
58
59
60

For peer review only

8. References

1. Stephenson, J.J., et al., *Clinical and economic burden of dyspnea and other COPD symptoms in a managed care setting*. Int J Chron Obstruct Pulmon Dis, 2017. **12**: p. 1947-1959.
2. David Ponka, M.K., *Top differential diagnoses in family medicine Dyspnoea Can Fam Physician*. Canadian Family Physician, 2008.

- 1
2
3 733 3. Figarska, S.M., H.M. Boezen, and J.M. Vonk, *Dyspnea severity, changes in dyspnea status and*
4 734 *mortality in the general population: the Vlagtwedde/Vlaardingen study*. European Journal of
5 735 Epidemiology, 2012. **27**(11): p. 867-876.
- 6 736 4. Parshall, M.B., et al., *An Official American Thoracic Society Statement: Update on the*
7 737 *Mechanisms, Assessment, and Management of Dyspnea*. American Journal of Respiratory and
8 738 Critical Care Medicine, 2012. **185**(4): p. 435-452.
- 9 739 5. Lim, W.S., et al., *BTS guidelines for the management of community acquired pneumonia in*
10 740 *adults: update 2009*. Thorax, 2009. **64 Suppl 3**: p. iii1-55.
- 11 741 6. Leuzzi, G., et al., *C-reactive protein level predicts mortality in COPD: a systematic review and*
12 742 *meta-analysis*. Eur Respir Rev, 2017. **26**(143).
- 13 743 7. Moe, G.W., *B-type natriuretic peptide in heart failure*. Curr Opin Cardiol, 2006. **21**(3): p. 208-14.
- 14 744 8. Ponikowski, P., et al., *2016 ESC Guidelines for the diagnosis and treatment of acute and chronic*
15 745 *heart failure: The Task Force for the diagnosis and treatment of acute and chronic heart failure*
16 746 *of the European Society of Cardiology (ESC) Developed with the special contribution of the Heart*
17 747 *Failure Association (HFA) of the ESC*. Eur Heart J, 2016. **37**(27): p. 2129-2200.
- 18 748 9. Alex Sakula, *Charcot leyden crystals and crushmann spirals*. Thorax, 1986.
- 19 749 10. Calhoun, W.J., J. Sedgwick, and W.W. Busse, *The role of eosinophils in the pathophysiology of*
20 750 *asthma*. Ann N Y Acad Sci, 1991. **629**: p. 62-72.
- 21 751 11. Pavord, I.D., *Eosinophilic phenotypes of airway disease*. Ann Am Thorac Soc, 2013. **10 Suppl**: p.
22 752 S143-9.
- 23 753 12. Tashkin, D.P. and M.E. Wechsler, *Role of eosinophils in airway inflammation of chronic*
24 754 *obstructive pulmonary disease*. International Journal of Chronic Obstructive Pulmonary Disease,
25 755 2018. **13**: p. 335-349.
- 26 756 13. Bafadhel, M., et al., *Blood Eosinophils to Direct Corticosteroid Treatment of Exacerbations of*
27 757 *Chronic Obstructive Pulmonary Disease: A Randomized Placebo-Controlled Trial*. American
28 758 Journal of Respiratory and Critical Care Medicine, 2012. **186**(1): p. 48-55.
- 29 759 14. Bedossa, P. and K. Patel, *Biopsy and Noninvasive Methods to Assess Progression of Nonalcoholic*
30 760 *Fatty Liver Disease*. Gastroenterology, 2016. **150**(8): p. 1811-1822 e4.
- 31 761 15. Fens, N., et al., *Exhaled breath analysis by electronic nose in airways disease. Established issues*
32 762 *and key questions*. Clin Exp Allergy, 2013. **43**(7): p. 705-15.
- 33 763 16. Probert, C.S., et al., *Volatile organic compounds as diagnostic biomarkers in gastrointestinal and*
34 764 *liver diseases*. J Gastrointestin Liver Dis, 2009. **18**(3): p. 337-43.
- 35 765 17. Schubert, J.K., et al., *Breath analysis in critically ill patients: potential and limitations*. Expert Rev
36 766 Mol Diagn, 2004. **4**(5): p. 619-29.
- 37 767 18. Fowler, S.J., et al., *Surveillance for lower airway pathogens in mechanically ventilated patients by*
38 768 *metabolomic analysis of exhaled breath: a case-control study*. Thorax, 2015. **70**(4): p. 320-5.
- 39 769 19. Florian Gahleitner, C.G.-H., Caroline S Beardsmore, Hitesh C Pandya, CL Paul Thomas,
40 770 *Metabolomics pilot study to identify VOC of childhood asthma in exhaled breath*. Bioanalysis,
41 771 2013.
- 42 772 20. Van Berkel, J.J., et al., *A profile of volatile organic compounds in breath discriminates COPD*
43 773 *patients from controls*. Respir Med, 2010. **104**(4): p. 557-63.
- 44 774 21. Basanta, M., et al., *Exhaled volatile organic compounds for phenotyping chronic obstructive*
45 775 *pulmonary disease: a cross-sectional study*. Respir Res, 2012. **13**: p. 72.
- 46 776 22. Besa, V., et al., *Exhaled volatile organic compounds discriminate patients with chronic*
47 777 *obstructive pulmonary disease from healthy subjects*. International Journal of Chronic
48 778 Obstructive Pulmonary Disease, 2015. **10**: p. 399-406.

- 1
2
3 779 23. Hauschild, A.C., J.I. Baumbach, and J. Baumbach, *Integrated statistical learning of metabolic ion*
4 780 *mobility spectrometry profiles for pulmonary disease identification*. Genet Mol Res, 2012. **11**(3):
5 781 p. 2733-44.
- 6 782 24. Jareno-Esteban, J.J., et al., *Study of 5 Volatile Organic Compounds in Exhaled Breath in Chronic*
7 783 *Obstructive Pulmonary Disease*. Archivos De Bronconeumologia, 2017. **53**(5): p. 251-256.
- 8 784 25. Paredi, P., et al., *Exhaled ethane, a marker of lipid peroxidation, is elevated in chronic obstructive*
9 785 *pulmonary disease*. Am J Respir Crit Care Med, 2000. **162**(2 Pt 1): p. 369-73.
- 10 786 26. Caldeira, M., et al., *Profiling allergic asthma volatile metabolic patterns using a headspace-solid*
11 787 *phase microextraction/gas chromatography based methodology*. J Chromatogr A, 2011.
12 788 **1218**(24): p. 3771-80.
- 13 789 27. Caldeira, M., et al., *Allergic asthma exhaled breath metabolome: a challenge for comprehensive*
14 790 *two-dimensional gas chromatography*. J Chromatogr A, 2012. **1254**: p. 87-97.
- 15 791 28. Dallinga, J.W., et al., *Volatile organic compounds in exhaled breath as a diagnostic tool for*
16 792 *asthma in children*. Clin Exp Allergy, 2010. **40**(1): p. 68-76.
- 17 793 29. Dragonieri, S., et al., *An electronic nose in the discrimination of patients with asthma and*
18 794 *controls*. J Allergy Clin Immunol, 2007. **120**(4): p. 856-62.
- 19 795 30. Fens, N., et al., *Exhaled breath profiling enables discrimination of chronic obstructive pulmonary*
20 796 *disease and asthma*. Am J Respir Crit Care Med, 2009. **180**(11): p. 1076-82.
- 21 797 31. Ibrahim, B., et al., *Non-invasive phenotyping using exhaled volatile organic compounds in*
22 798 *asthma*. Thorax, 2011. **66**(9): p. 804-9.
- 23 799 32. Montuschi, P., et al., *Diagnostic performance of an electronic nose, fractional exhaled nitric*
24 800 *oxide, and lung function testing in asthma*. Chest, 2010. **137**(4): p. 790-6.
- 25 801 33. Christiansen, A., et al., *A systematic review of breath analysis and detection of volatile organic*
26 802 *compounds in COPD*. J Breath Res, 2016. **10**(3): p. 034002.
- 27 803 34. de Vries, R., et al., *Integration of electronic nose technology with spirometry: validation of a new*
28 804 *approach for exhaled breath analysis*. J Breath Res, 2015. **9**(4): p. 046001.
- 29 805 35. Shafiek, H., et al., *Using the Electronic Nose to Identify Airway Infection during COPD*
30 806 *Exacerbations*. PLoS One, 2015. **10**(9): p. e0135199.
- 31 807 36. Finamore, P., et al., *Analysis of volatile organic compounds: an innovative approach to heart*
32 808 *failure characterization in older patients*. J Breath Res, 2018. **12**(2): p. 026007.
- 33 809 37. *MRC Molecular Pathology, Review 20 Aug*.
- 34 810 38. Medical Research Council, *Molecular Pathology Nodes*. 2014: MRC website.
- 35 811 39. Corradi, M., et al., *Aldehydes in exhaled breath condensate of patients with chronic obstructive*
36 812 *pulmonary disease*. Am J Respir Crit Care Med, 2003. **167**(10): p. 1380-6.
- 37 813 40. Larstad, M., et al., *Determination of malondialdehyde in breath condensate by high-performance*
38 814 *liquid chromatography with fluorescence detection*. J Chromatogr B Analyt Technol Biomed Life
39 815 *Sci*, 2002. **766**(1): p. 107-14.
- 40 816 41. F. Y. Hsieh, D.A.B., Michael D. Larsen,, *A simple method of sample size calculation for linear and*
41 817 *logistic regression* Statis Med, 1998.
- 42 818 42. Hajian-Tilaki, K., *Sample size estimation in diagnostic test studies of biomedical informatics*. J
43 819 *Biomed Inform*, 2014. **48**: p. 193-204.
- 44 820 43. Quan, H., et al., *Updating and validating the Charlson comorbidity index and score for risk*
45 821 *adjustment in hospital discharge abstracts using data from 6 countries*. Am J Epidemiol, 2011.
46 822 **173**(6): p. 676-82.
- 47 823 44. Roffman, C.E., J. Buchanan, and G.T. Allison, *Charlson Comorbidities Index*. J Physiother, 2016.
48 824 **62**(3): p. 171.
- 49 825 45. Bjarnason, H. and P. Hougaard, *Fisher information for two gamma frailty bivariate Weibull*
50 826 *models*. Lifetime Data Anal, 2000. **6**(1): p. 59-71.

- 1
2
3 827 46. Lawrence Gould, A., et al., *Joint modeling of survival and longitudinal non-survival data: current methods and issues. Report of the DIA Bayesian joint modeling working group.* Stat Med, 2015.
4 828
5 829 **34**(14): p. 2181-95.
- 6 830 47. Steer, J., et al., *Dyspnoea severity and pneumonia as predictors of in-hospital mortality and early readmission in acute exacerbations of COPD.* Thorax, 2012. **67**(2): p. 117-21.
7 831
- 8 832 48. Tsimogianni, A.M., et al., *Predictors of Outcome After Exacerbation of Chronic Obstructive Pulmonary Disease.* Journal of General Internal Medicine, 2009. **24**(9): p. 1043-1048.
9 833
- 10 834 49. Kitchen, S., et al., *LATE-BREAKING ABSTRACT: Breathe free: Open source development of a breath sampler by a consortium of breath researchers.* European Respiratory Journal, 2015.
11 835
12 836 **46**(suppl 59).
- 13 837 50. Oostveen, E., et al., *The forced oscillation technique in clinical practice: methodology, recommendations and future developments.* Eur Respir J, 2003. **22**(6): p. 1026-41.
14 838
- 15 839 51. Pendharkar, S. and S. Mehta, *The clinical significance of exhaled nitric oxide in asthma.* Canadian Respiratory Journal : Journal of the Canadian Thoracic Society, 2008. **15**(2): p. 99-106.
16 840
- 17 841 52. Yates, D.H., *Role of exhaled nitric oxide in asthma.* Immunol Cell Biol, 2001. **79**(2): p. 178-90.
18 842
- 19 843 53. Lang, R.M., et al., *Recommendations for chamber quantification: a report from the American Society of Echocardiography's Guidelines and Standards Committee and the Chamber Quantification Writing Group, developed in conjunction with the European Association of Echocardiography, a branch of the European Society of Cardiology.* J Am Soc Echocardiogr, 2005.
20 844
21 845 **18**(12): p. 1440-63.
- 22 846 54. J. Throck Watson, O.D.S., *Introduction to Mass Spectrometry: Instrumentation, Applications, and Strategies for Data Interpretation, 4th Edition.* 2007: Wiley. 862.
23 847
24 848
- 25 849 55. Buszewski, B., et al., *Human exhaled air analytics: biomarkers of diseases.* Biomed Chromatogr, 2007. **21**(6): p. 553-66.
26 850
- 27 851 56. Amann, A., et al., *Applications of breath gas analysis in medicine.* International Journal of Mass Spectrometry, 2004. **239**(2-3): p. 227-233.
28 852
- 29 853 57. Almstetter, M.F., P.J. Oefner, and K. Dettmer, *Comprehensive two-dimensional gas chromatography in metabolomics.* Anal Bioanal Chem, 2012. **402**(6): p. 1993-2013.
30 854
- 31 855 58. Das, M.K., et al., *Investigation of Gender-Specific Exhaled Breath Volatome in Humans by GCxGC-TOF-MS.* Analytical Chemistry, 2014. **86**(2): p. 1229-1237.
32 856
- 33 857 59. Phillips, M., et al., *Detection of an extended human volatome with comprehensive two-dimensional gas chromatography time-of-flight mass spectrometry.* PLoS One, 2013. **8**(9): p. e75274.
34 858
35 859
- 36 860 60. Gruber, B., et al., *Vacuum ultraviolet absorption spectroscopy in combination with comprehensive two-dimensional gas chromatography for the monitoring of volatile organic compounds in breath gas: A feasibility study.* J Chromatogr A, 2016. **1464**: p. 141-6.
37 861
38 862
- 39 863 61. Gruber, B., et al., *Breath gas monitoring during a glucose challenge by a combined PTR-QMS/GCxGC-TOFMS approach for the verification of potential volatile biomarkers.* J Breath Res, 2016. **10**(3): p. 036003.
40 864
41 865
- 42 866 62. Mellors, T.R., et al., *A new method to evaluate macaque health using exhaled breath: A case study of M. tuberculosis in a BSL-3 setting.* J Appl Physiol (1985), 2017. **122**(3): p. 695-701.
43 867
- 44 868 63. Phillips, M., et al., *Breath biomarkers of whole-body gamma irradiation in the Gottingen minipig.* Health Phys, 2015. **108**(5): p. 538-46.
45 869
- 46 870 64. Bajtarevic, A., et al., *Noninvasive detection of lung cancer by analysis of exhaled breath.* BMC Cancer, 2009. **9**: p. 348.
47 871
- 48 872 65. Lavra, L., et al., *Investigation of VOCs associated with different characteristics of breast cancer cells.* Sci Rep, 2015. **5**: p. 13246.
49 873
50
51
52
53
54
55
56
57
58
59
60

- 1
2
3 874 66. Zhou, W., et al., *Exhaled breath online measurement for cervical cancer patients and healthy*
4 875 *subjects by proton transfer reaction mass spectrometry*. *Anal Bioanal Chem*, 2017. **409**(23): p.
5 876 5603-5612.
- 6 877 67. Fernández del Río, R., et al., *Volatile Biomarkers in Breath Associated With Liver Cirrhosis —*
7 878 *Comparisons of Pre- and Post-liver Transplant Breath Samples*. *EBioMedicine*, 2015. **2**(9): p.
8 879 1243-1250.
- 10 880 68. O'Hara, M.E., et al., *Limonene in exhaled breath is elevated in hepatic encephalopathy*. *J Breath*
11 881 *Res*, 2016. **10**(4): p. 046010.
- 12 882 69. Tomasz Ligor, C.A., Anton Amann, Konrad Schwarz, Bogusaw Buszewski, Wojciech ebrowski,,
13 883 *<Comparison of Proton Transfer Reaction-Mass Spectrometry and Gas Chromatography-Mass*
14 884 *Spectrometry in Analysis of Breath Samples.pdf>*. *Chem. Anal.*, 2009.
- 15 885 70. Darwiche, K., et al., *Bronchoscopically obtained volatile biomarkers in lung cancer*. *Lung*, 2011.
16 886 **189**(6): p. 445-52.
- 18 887 71. Handa, H., et al., *Exhaled Breath Analysis for Lung Cancer Detection Using Ion Mobility*
19 888 *Spectrometry*. *PLoS ONE*, 2014. **9**(12): p. e114555.
- 20 889 72. Bessa, V., et al., *Detection of volatile organic compounds (VOCs) in exhaled breath of patients*
21 890 *with chronic obstructive pulmonary disease (COPD) by ion mobility spectrometry*. *International*
22 891 *Journal for Ion Mobility Spectrometry*, 2011. **14**(1): p. 7-13.
- 23 892 73. Schivo, M., et al., *Mobile instrumentation platform to distinguish airway disorders*. *Journal of*
24 893 *breath research*, 2013. **7**(1): p. 017113-017113.
- 26 894 74. Badjagbo, K., et al., *Direct atmospheric pressure chemical ionization-tandem mass spectrometry*
27 895 *for the continuous real-time trace analysis of benzene, toluene, ethylbenzene, and xylenes in*
28 896 *ambient air*. *J Am Soc Mass Spectrom*, 2009. **20**(5): p. 829-36.
- 29 897 75. Healy, R.M., et al., *A multi-year study of VOC emissions at a chemical waste disposal facility*
30 898 *using mobile APCI-MS and LPCI-MS instruments*. *Environ Pollut*, 2018. **232**: p. 220-228.
- 31 899 76. Linforth, R. and A.J. Taylor, *Persistence of Volatile Compounds in the Breath after Their*
32 900 *Consumption in Aqueous Solutions*. *Journal of Agricultural and Food Chemistry*, 2000. **48**(11): p.
33 901 5419-5423.
- 34 902 77. Perez-Hurtado, P., et al., *Direct analysis of volatile organic compounds in foods by headspace*
35 903 *extraction atmospheric pressure chemical ionisation mass spectrometry*. *Rapid Commun Mass*
36 904 *Spectrom*, 2017. **31**(22): p. 1947-1956.
- 38 905 78. Berchtold, C., L. Meier, and R. Zenobi, *Evaluation of extractive electrospray ionization and*
39 906 *atmospheric pressure chemical ionization for the detection of narcotics in breath*. *International*
40 907 *Journal of Mass Spectrometry*, 2011. **299**(2-3): p. 145-150.
- 41 908 79. Kapishon, V., et al., *Atmospheric pressure chemical ionization mass spectrometry of pyridine and*
42 909 *isoprene: potential breath exposure and disease biomarkers*. *J Breath Res*, 2013. **7**(2): p. 026005.
- 43 910 80. Heaney, L.M., et al., *Real-time monitoring of exhaled volatiles using atmospheric pressure*
44 911 *chemical ionization on a compact mass spectrometer*. *Bioanalysis*, 2016. **8**(13): p. 1325-36.
- 45 912 81. Lê Cao, K.-A., S. Boitard, and P. Besse, *Sparse PLS discriminant analysis: biologically relevant*
46 913 *feature selection and graphical displays for multiclass problems*. *BMC Bioinformatics*, 2011. **12**:
47 914 p. 253-253.
- 49 915 82. Zou, H., T. Hastie, and R. Tibshirani, *Sparse Principal Component Analysis*. *Journal of*
50 916 *Computational and Graphical Statistics*, 2006. **15**(2): p. 265-286.
- 51 917 83. Zou H, H.T., *Regularization and Variable Selection via the Elastic Net*. *Journal of the Royal*
52 918 *Statistical Society*, 2005.

919

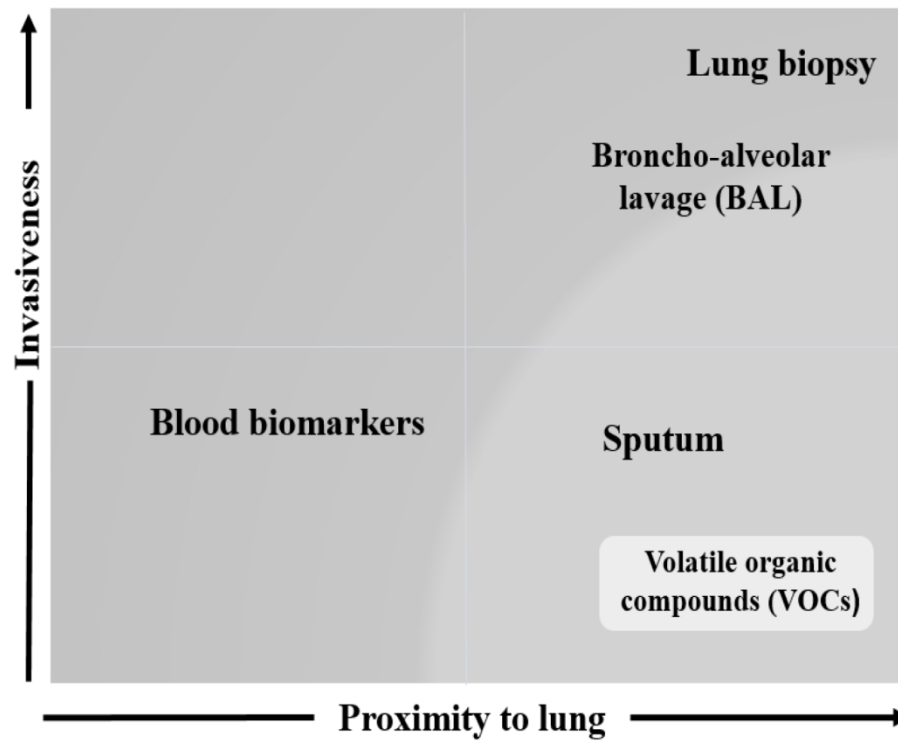


Figure (1): Relationship between lung proximity and degree of invasiveness of different lung matrices. The figure plots the level of invasiveness of various lung matrices in relation to their proximity to the lung. Given their pathological relevance, the degree of invasiveness of bronchoalveolar lavage (BAL) and lung biopsy makes them less favourable in diagnosing respiratory diseases.

144x115mm (300 x 300 DPI)

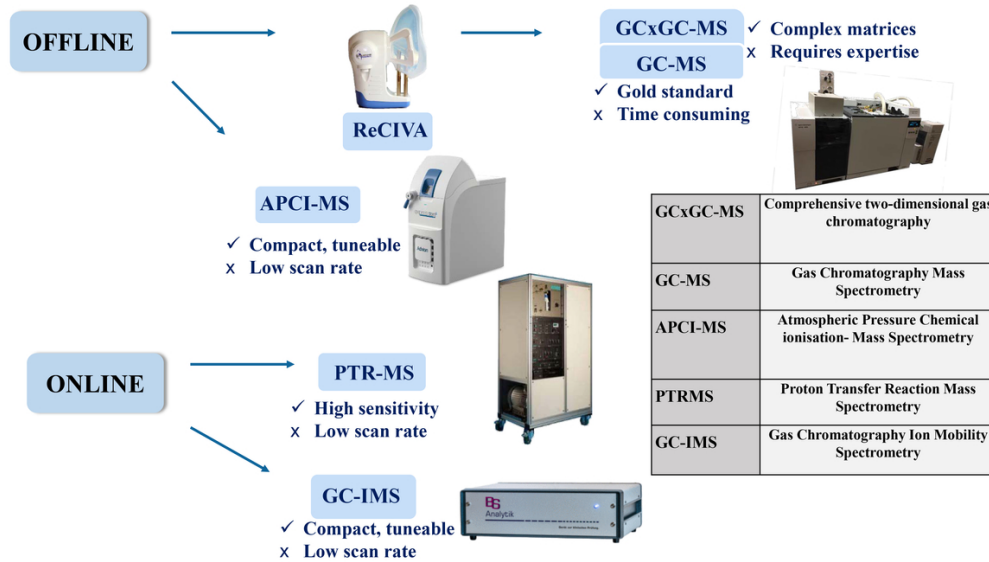


Figure (2): Multi-instrument use in breath sampling. Figure illustrates the various combinations of offline and online devices used in breath sampling and the relevant pros and cons. Offline and online technologies are used for the discovery and validation phases of the study respectively.

97x55mm (300 x 300 DPI)

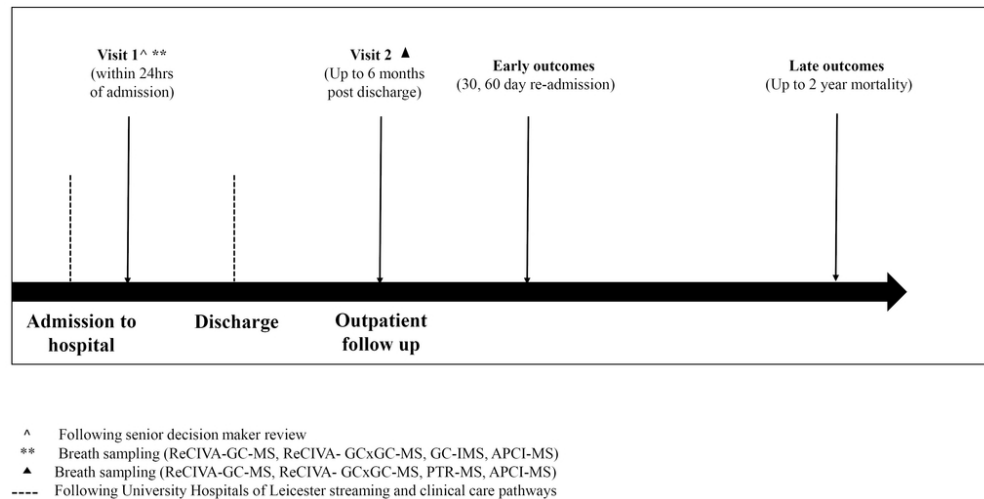


Figure (3): Study flow chart. Figure outlines the patient journey from admission through to discharge and follow up. Participants with self-reported acute breathlessness presenting to University Hospitals of Leicester are recruited within 24 hours following a senior decision maker review. Breath sampling is carried out on the first visit, at recruitment, and the second visit, up to 6 months after discharge. Patients are admitted through the standard operational emergency medical streaming and care pathways at the University Hospitals of Leicester NHS Trust. Early outcomes (hospital readmission) are measured at 30 and 60 days and late outcomes (mortality) including respiratory and all-cause mortality are measured at 2 years. Assessments carried out at each time point are summarised in Table 1.

86x43mm (300 x 300 DPI)

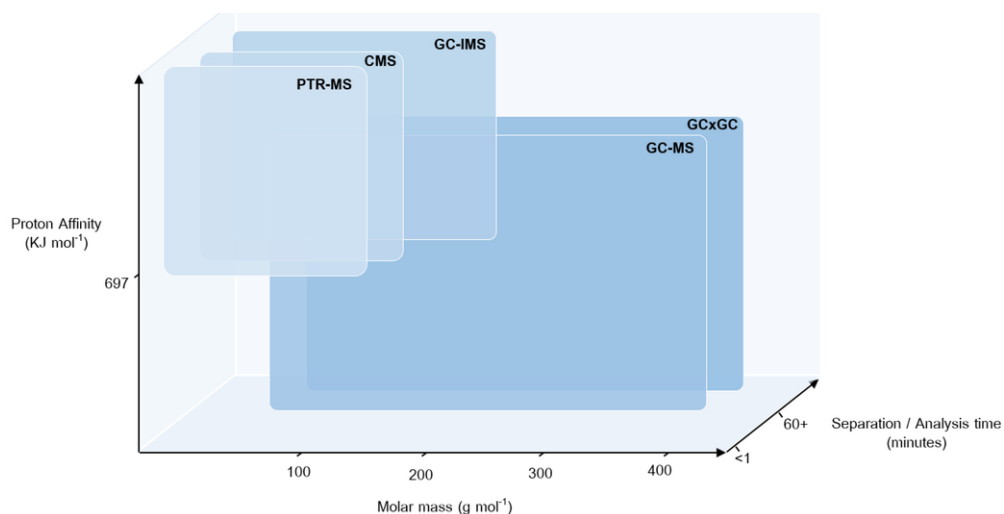


Figure (4): Multi-instrument use in breath sampling. Operational space of the analytical technologies used in EMBER for the analysis of volatile organic compounds (VOC) in exhaled breath, including proton transfer reaction-mass spectrometry (PTR-MS), atmospheric pressure chemical ionisation-mass spectrometry (CMS), gas chromatography-ion mobility spectrometry (GC-IMS), gas chromatography-mass spectrometry (GC-MS) and two-dimensional gas chromatography-mass spectrometry (GCxGC). Comparing the typical molar mass range detectable; selectivity in detection owing to the type of ionisation involved and the proton affinity of analytes; and the inclusion of a chromatographic separation affecting total time of analysis. The online technologies involving chemical ionisation (PTR-MS, CMS and GC-IMS) can be used in-clinic owing to short analysis times, but only detect lower molar mass molecules with a proton affinity higher than 697 KJ mol⁻¹. Offline chromatographic techniques (GC-MS and GCxGC) detect a wider range of compounds independent of proton affinity; however, the techniques have longer analysis times and involve sample transportation and storage.

88x44mm (300 x 300 DPI)

BMJ Open

Assessment of Breath Volatile Organic Compounds in Acute Cardio-respiratory Breathlessness: A protocol describing a Prospective Real World Observational Study

Journal:	<i>BMJ Open</i>
Manuscript ID	bmjopen-2018-025486.R2
Article Type:	Protocol
Date Submitted by the Author:	23-Nov-2018
Complete List of Authors:	<p>Ibrahim, Wadah; Leicester NIHR Biomedical Research Centre (Respiratory theme); College of Life Sciences, Department of Infection, Immunity and Inflammation, University of Leicester</p> <p>Wilde, Michael; University of Leicester, Department of Chemistry</p> <p>Cordell, Rebecca; University of Leicester, Department of Chemistry</p> <p>Salman, Dahlia; Loughborough University, Department of Chemistry</p> <p>Ruszkiewicz, Dorota; Loughborough University, Department of Chemistry</p> <p>Bryant, Luke; University of Leicester, Department of Chemistry</p> <p>Richardson, Matthew; Leicester NIHR Biomedical Research Centre (Respiratory theme); University of Leicester, Infection, Inflammation and Immunity</p> <p>Free, Robert; Leicester NIHR Biomedical Research Centre (Respiratory theme); University of Leicester, Infection, Inflammation and Immunity</p> <p>Zhao, Bo; Leicester NIHR Biomedical Research Centre (Respiratory theme); University of Leicester, Infection, Inflammation and Immunity</p> <p>Yousuf, Ahmed; University of Leicester, Infection, Inflammation and Immunity; Leicester NIHR Biomedical Research Centre (Respiratory theme)</p> <p>White, Christobelle; University of Leicester, Infection, Inflammation and Immunity; Leicester NIHR Biomedical Research Centre (Respiratory theme)</p> <p>Russell, Richard; University of Leicester, Infection, Inflammation and Immunity; Leicester NIHR Biomedical Research Centre (Respiratory theme)</p> <p>Jones, Sheila; University Hospitals of Leicester NHS Trust</p> <p>Patel, Bharti; University Hospitals of Leicester NHS Trust</p> <p>Awal, Asia; University Hospitals of Leicester NHS Trust</p> <p>Phillips, Rachael; NIHR Leicester Clinical Research Facility</p> <p>Fowkes, Graham; NIHR Leicester Clinical Research Facility</p> <p>McNally, Teresa; Paediatric Clinical Investigation Centre</p> <p>Foxon, Clare; Paediatric Clinical Investigation Centre</p> <p>Bhatt, Hetan; Leicester NIHR Biomedical Research Centre (Respiratory theme)</p> <p>Peltrini, Rosa; Leicester NIHR Biomedical Research Centre (Respiratory theme); University of Leicester, Department Infection, Inflammation and Immunity</p> <p>Singapuri, Amisha; Leicester NIHR Biomedical Research Centre (Respiratory theme)</p> <p>Hargadon, Beverley; Leicester NIHR Biomedical Research Centre</p>

	(Respiratory theme); University of Leicester, Infection, Inflammation and Immunity Suzuki, T; Department of Cardiovascular Sciences, University of Leicester, , Cardiovascular Research Centre; Leicester NIHR Biomedical Research Centre (Cardiovascular theme) Ng, Leong; Department of Cardiovascular Sciences, University of Leicester, Cardiovascular Research Centre; Leicester NIHR Biomedical Research Centre (Cardiovascular theme) Gaillard, Erol; Leicester NIHR Biomedical Research Centre (Respiratory theme); Paediatric Clinical Investigation Centre Beardsmore, Caroline; Leicester NIHR Biomedical Research Centre (Respiratory theme); Paediatric Clinical Investigation Centre Ryanna, Kimuli; University Hospitals of Leicester NHS Trust Pandya, Hitesh; Discovery Medicine, Respiratory Therapeutic Area, GlaxoSmithKline PLC Coates, Tim; University of Leicester, Infection, Inflammation and Immunity; Department of Cardiovascular Sciences, University of Leicester Monks, Paul; University of Leicester, Infection, Inflammation and Immunity; University of Leicester, Department of Chemistry Greening, Neil; Leicester NIHR Biomedical Research Centre (Respiratory theme); University of Leicester, Infection, Inflammation and Immunity Brightling, Christopher; Leicester NIHR Biomedical Research Centre (Respiratory theme); University of Leicester, Infection, Inflammation and Immunity Thomas, Paul; Loughborough University, Department of Chemistry Siddiqui, Salman; Leicester NIHR Biomedical Research Centre (Respiratory theme); University of Leicester, Infection, Inflammation and Immunity
Primary Subject Heading:	Respiratory medicine
Secondary Subject Heading:	Research methods, Respiratory medicine
Keywords:	Breathlessness, Observational study, Volatile Organic Compound, Breath analysis

SCHOLARONE™
Manuscripts

1 **Assessment of Breath Volatile Organic Compounds in Acute Cardio-** 2 **respiratory Breathlessness: A protocol describing a Prospective Real** 3 **World Observational Study**

4
 5 Wadah Ibrahim^{1,2}, Michael Wilde³, Rebecca Cordell³, Dahlia Salman⁴, Dorota Ruszkiewicz⁴, Luke
 6 Bryant³, Matthew Richardson^{1,2}, Robert C. Free^{1,2}, Bo Zhao^{1,2}, Ahmed Yousuf^{1,2}, Christobelle
 7 White^{1,2}, Richard J Russell^{1,2}, Sheila Jones⁵, Bharti Patel⁵, Asia Awal⁵, Rachael Phillips⁶, Graham
 8 Fowkes⁶, Teresa McNally⁷, Clare Foxon⁷, Hetan Bhatt^{1,2}, Rosa Peltrini^{1,2}, Amisha Singapuri^{1,2},
 9 Beverley Hargadon^{1,2}, Toru Suzuki^{8,9}, Leong L. Ng^{8,9}, Erol A Gaillard^{2,7}, Caroline Beardsmore^{2,7},
 10 Kimuli Ryanna⁵, Hitesh Pandya¹¹, Tim Coates^{2,10}, Paul S. Monks^{2,3}, Neil Greening^{1,2}, Chris E
 11 Brightling^{1,2}, Paul Thomas⁴, Salman Siddiqui^{1,2}

12
 13
 14 wadah.ibrahim@leicester.ac.uk; mjw77@leicester.ac.uk; rc145@leicester.ac.uk; d.salman@lboro.ac.uk;
 15 d.m.ruszkiewicz2@lboro.ac.uk; lb368@leicester.ac.uk; mr251@leicester.ac.uk; rcf8@leicester.ac.uk;
 16 bo.zhao@leicester.ac.uk; ajy7@leicester.ac.uk; cmjw3@leicester.ac.uk; rjr22@leicester.ac.uk;
 17 rjr22@leicester.ac.uk; sheila.jones@uhl-tr.nhs.uk; bharti.m.patel@uhl-tr.nhs.uk; asia.awal@uhl-tr.nhs.uk;
 18 rachael.phillips@uhl-tr.nhs.uk; graham.fowkes@uhl-tr.nhs.uk; tm23@leicester.ac.uk; clare.foxon@uhl-
 19 tr.nhs.uk; hetan.bhatt@uhl-tr.nhs.uk; rp417@leicester.ac.uk; amisha.singapuri@uhl-tr.nhs.uk;
 20 beverley.hargadon@uhl-tr.nhs.uk; ts263@le.ac.uk; lln1@le.ac.uk; eag15@leicester.ac.uk; csb@leicester.ac.uk;
 21 kimuli.ryanna@uhl-tr.nhs.uk; hitesh.x.pandya@gsk.com; tim.coates@leicester.ac.uk;
 22 paul.monks@leicester.ac.uk; neil.greening@leicester.ac.uk; ceb17@leicester.ac.uk;
 23 C.L.P.Thomas@lboro.ac.uk; ss338@leicester.ac.uk

24
 25 Corresponding author: Professor Salman Siddiqui, ss338@leicester.ac.uk

26 NIHR Leicester Biomedical Research Centre (Respiratory Theme)

27 Glenfield Hospital,

28 Groby Road

29 Leicester

30 LE3 9QP

31 t: +44 116 256 3841

32 e: ss338@leicester.ac.uk

33 34 **Authors' affiliations:**

35 ¹Leicester NIHR Biomedical Research Centre (Respiratory theme), Glenfield Hospital, Groby Road, Leicester
 36 LE3 9QP

37 ² College of Life Sciences, Department of Infection, Immunity and Inflammation, University of Leicester,
 38 University Road, Leicester LE1 7RH

- 1
2
3 39 ³ Department of Chemistry, University of Leicester, University Road, Leicester, LE1 7RH
4
5 40 ⁴ Department of Chemistry, Loughborough University, Epinal Way, Loughborough, LE11 3TU
6
7 41 ⁵ University Hospitals of Leicester NHS Trust, Glenfield Hospital, Groby Road, Leicester LE3 9QP
8
9 42 ⁶ NIHR Leicester Clinical Research Facility, Research Space, Level 0, Balmoral Building, Leicester Royal
10 Infirmery, Infirmery Square, Leicester, LE1 5WW
11
12 43
13 44 ⁷ Paediatric Clinical Investigation Centre, Leicester Respiratory Biomedical Research Centre, University of
14 Leicester, Leicester Royal Infirmery, Leicester, LE2 7LX
15
16 45
17 46 ⁸ Department of Cardiovascular Sciences, University of Leicester, Cardiovascular Research Centre, Glenfield
18 General Hospital, Leicester LE3 9QP.
19
20 47
21 48 ⁹ Leicester NIHR Biomedical Research Centre (Cardiovascular theme), Glenfield Hospital, Groby Road,
22 Leicester LE3 9QP
23
24 49
25 50 ¹⁰ Department of Cardiovascular Sciences, University of Leicester, Level G, Jarvis building RMO, Infirmery
26 Square, Leicester LE1 5WW
27
28 51
29 52 ¹¹ Discovery Medicine, Respiratory Therapeutic Area, GlaxoSmithKline PLC, Gunnels Wood Road, Stevenage,
30 Hertfordshire, SG1 2NY
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

review only

69

Abstract**Introduction**

Patients presenting with acute undifferentiated breathlessness are commonly encountered in admissions units across the United Kingdom. Existing blood biomarkers have clinical utility in distinguishing patients with single organ pathologies but have poor discriminatory power in multi factorial presentations. Evaluation of volatile organic compounds (VOC) in exhaled breath offers the potential to develop biomarkers of disease states that underpin acute cardio-respiratory breathlessness, owing to their proximity to the cardio-respiratory system. To date there has been no systematic evaluation of VOC in acute cardio-respiratory breathlessness. The proposed study will seek to use both offline and online VOC technologies to evaluate the predictive value of VOC in identifying common conditions that present with acute cardio-respiratory breathlessness.

Methods and analysis

A prospective real world observational study carried out across three acute admissions units within Leicestershire. Participants with self-reported acute breathlessness, with a confirmed primary diagnosis of either acute heart failure, community acquired pneumonia and acute exacerbation of asthma or COPD will be recruited within 24 hours of admission. Additionally, school age children admitted with severe asthma will be evaluated. All participants will undergo breath sampling on admission and upon recovery following discharge. A range of online technologies including: proton-transfer-reaction mass spectrometry (PTR-MS), gas chromatography ion mobility spectrometry (GC-IMS), atmospheric pressure chemical ionisation- mass spectrometry (APCI-MS) and offline technologies including gas chromatography mass spectroscopy (GC-MS) and comprehensive two-dimensional gas chromatography-mass spectrometry (GCxGC-MS) will be utilised for VOC discovery and replication. For offline technologies a standardised CE marked breath sampling device (ReCIVA[®]) will be used. All recruited participants will be characterised using existing blood biomarkers including C - reactive protein (CRP), brain derived natriuretic peptide (BNP), Troponin-I

1
2
3 95 and blood eosinophil levels and further evaluated using a range of standardised questionnaires, lung
4
5 96 function testing, sputum cell counts and other diagnostic tests pertinent to acute disease.
6
7

8 97 **Ethics and dissemination**

9
10 98 The National Research Ethics Service Committee East Midlands has approved the study protocol
11
12 99 (REC number: 16/LO/1747). IRAS 198921. Findings will be presented at academic conferences and
13
14 100 published in peer-reviewed scientific journals. Dissemination will be facilitated via a partnership with
15
16 101 the East Midlands Academic Health Sciences Network and via interaction with all UK funded
17
18 102 MRC/EPSRC molecular pathology nodes.
19
20
21
22 103
23
24

25 104 **Key words:** Breathlessness; Breath analysis; Volatile Organic Compound, Observational study
26
27 105
28
29
30 106
31
32
33 107
34
35
36 108
37
38
39 109
40
41
42 110
43
44
45 111
46
47
48 112
49
50
51
52 113
53
54
55 114
56
57
58 115
59
60

116 **Strengths and Limitations of this Study**

- 117 • A pragmatic real world, prospective, observational study across three admission units that
118 focuses on the systematic discovery and replication of VOC in acutely breathless patients
119 using both online and offline technologies
- 120 • The proposed study is the largest of its kind in acute disease to characterise VOC with a
121 range of additional assessments that will build a comprehensive phenotype of acute
122 cardio-respiratory exacerbations
- 123 • The proposed study will build an infrastructure for research and subsequent evaluation of
124 VOC in interventional trials within acute cardio-respiratory exacerbations
- 125 • Prior acute treatment exposure will need to be accounted for when evaluating potential
126 discriminative biomarkers
- 127 • VOC technologies are not currently suited for deployment in patients that are of high
128 clinical acuity

130 **1. Introduction:**

131 Breathlessness is a common symptom of cardio-respiratory illnesses that has a significant direct
132 impact on patients' wellbeing as well as a substantial economic burden on healthcare systems [1].

133 Although its etiologies can be variable, exacerbations of common complex chronic cardio-respiratory
134 conditions account for approximately 70% of acute presentations with breathlessness, namely
135 exacerbations of asthma and COPD, acute heart failure and community acquired pneumonia [2].

136 Moreover, moderate and severe breathlessness is significantly associated with all-cause,
137 cardiovascular and COPD mortality[3]. As a consequence symptomatic breathlessness warrants rapid
138 evaluation and targeted diagnostics at presentation.

139
140 Diagnostic evaluation of acute breathlessness is heavily reliant on blood based biomarkers e.g. CRP,
141 BNP, Troponin and on occasions blood eosinophil levels. These biomarkers have clinical utility
142 primarily in patients with single pathologies, but have poor discriminatory power in patients with

1
2
3 143 multifactorial presentations of acute breathlessness[4]. There is therefore an unmet need for the
4
5 144 development of sensitive and specific biomarkers that differentiate acute breathlessness from its
6
7 145 recovery and the common cardio-respiratory conditions that present with acute breathlessness.
8
9 146
10
11 147 CRP plays an important role in diagnosing breathlessness caused by an underlying bacterial
12
13 148 pneumonia[5], as well as predicting mortality in patients with COPD [6]. BNP is routinely utilised in
14
15 149 acute settings to support the diagnosis of acute heart failure [7]. The European Society of Cardiology
16
17 150 (ESC) recommends BNP threshold values of <100 pg/mL to rule out acute congestive cardiac failure
18
19 151 and values > 500 pg/ml as diagnostic of acute exacerbations of heart failure [8].
20
21
22 152

23
24 153 The role of peripheral blood eosinophil count in airway inflammation was poorly understood up until
25
26 154 the second half of the 19th century when Paul Ehrlich, a German physician and Nobel prize winner,
27
28 155 introduced eosin in his technique for white cell differentiation in 1879[9]. Considerable advances in
29
30 156 the field of airway inflammation and the role of eosinophils have taken place since [10-12]. More
31
32 157 recently Bafadhel *et al* suggested that peripheral blood eosinophil count can be used to direct
33
34 158 corticosteroid therapy during COPD exacerbations in single centre study [13].
35
36

37 159 Currently, blood biomarkers together with clinical, physiological and imaging parameters are used in
38
39 160 diagnosing the cause of acute breathlessness. Blood biomarkers may be less specific as they originate
40
41 161 far from the target organs of interest (the heart and the lungs in cardio-respiratory disease). Sputum,
42
43 162 although potentially a more definitive lung specific matrix, is comparatively difficult to obtain
44
45 163 particularly in acutely unwell patients, limiting its use in acute disease and highlighting the need for
46
47 164 better biomarkers. Ideally these biomarkers would have the following characteristics, (i) they would
48
49 165 originate from the target organ of interest, (ii) they would significantly add value to conventional risk
50
51 166 scoring and diagnostic algorithms in acute breathlessness, (iii) they would be minimally invasive and
52
53 167 suitable for rapid point of care diagnosis in emergency rooms and acute admissions units (iv) they
54
55 168 would have diagnostic value in patients with multifactorial acute breathlessness.
56
57
58 169

1
2
3 170 Exhaled breath contains thousands of volatile organic compounds (VOC) that reflect biological
4
5 171 processes occurring in the host both locally in the airways and systematically offering the potential to
6
7 172 develop more effective biomarkers in acutely breathless patients (**Figure 1**).

8
9 173
10
11 174 The proposed program of research will use a combination of offline and online technologies to
12
13 175 identify and evaluate the diagnostic and prognostic value of VOC in patients with acute cardio-
14
15 176 respiratory related breathlessness (**Figure 2**).

16
17 177
18
19
20 178 Exhaled breath analysis of volatile chemicals offers the potential to develop more effective biomarkers
21
22 179 in acutely breathless patients. The use of breath analysis for disease diagnostics dates back to ancient
23
24 180 Greeks where physicians used exhaled breath to diagnose different diseases. Breath odours allow
25
26 181 correct associations to certain diseases. For example, the sweet smell of diabetic ketoacidosis, the
27
28 182 fishy smell of breath associated to liver illness, the urine-like odour of kidney disease and the smell of
29
30 183 the breath of patients with lung abscesses, caused by the proliferation of anaerobic bacteria [14-17].
31
32 184 More recently, exhaled breath analysis has demonstrated early proof of concept in the diagnosis of
33
34 185 acute heart failure, and ventilator associated pneumonia[18] . The validity of breath analysis has also
35
36 186 been demonstrated in breathless children[19]. This population is likely to prefer breath-based tests, as
37
38 187 these are minimally invasive. Importantly, a variety of point of care sensors are now available to
39
40 188 evaluate potential exhaled breath biomarkers in emergency care settings.

41
42 189
43
44
45 190 A study by Van Berkel et al demonstrated the ability to distinguish COPD subjects from controls
46
47 191 solely based on the presence of VOCs in breath, suggesting that analysis of VOC might be highly
48
49 192 relevant for diagnosis of COPD [20]. This established the basis of further studies of VOC in COPD
50
51 193 [21-25].recommending larger studies for validation.

52
53 194 Several other studies found that VOC profiling in diagnosing asthma is potentially feasible [26-32].
54
55 195 This however has been done in relatively small numbers in stable disease.

56
57 196 Despite the novelty of non-invasive sampling technology and the growing interest in exhaled breath
58
59 197 analysis there remains a disappointing level of comparability across studies due to the lack of

1
2
3 198 standardisation and appropriate data analysis methods. A recent systemic review by Anders
4
5 199 Christiansen *et al* compared eleven publications reporting very heterogeneous designs, methods,
6
7 200 patient group sizes, data analytics and, consequently, quite varying results [33].
8

9 201
10
11 202 To our knowledge, no other large studies exploring the use of breath biomarkers in profiling acute
12
13 203 breathlessness have been completed. Several studies have explored the use of electronic nose (eNose)
14
15 204 in stable disease with good discriminatory power in COPD [34], Pneumonia [35] and heart failure[36]
16
17 205 with relatively small sample size. While eNose has now been widely used in detecting various VOC
18
19 206 patterns, GC-MS, a largely validated methodology, remains the gold standard technique for detecting
20
21 207 VOCs in exhaled breath. The focus of the current research study will be to evaluate acutely breathless
22
23 208 cardio-respiratory patients using a combination of ‘discovery’ and near-patient care breath sampling
24
25 209 technologies.

26
27
28 210 Medical Research Council (MRC) and Engineering and Physical Sciences Research Council (EPSRC)
29
30 211 have commissioned a series of molecular pathology nodes aimed at developing molecular signatures
31
32 212 relevant to disease diagnosis and progression. This was triggered by the clear need for alliance
33
34 213 between academic institutions, industry and NHS partners to enhance the benefits of stratified
35
36 214 medicine for patients[37, 38].

37
38
39 215 University of Leicester and Loughborough University were awarded a joint molecular pathology node
40
41 216 East Midlands Breathomics Pathology Node (EMBER) which this study forms a key part of.

42
43 217

44 45 218 **2. Methods and Analysis**

46 47 219 **2.1. Study design**

48
49 220 A prospective real world observational study across three acute admissions units within Leicestershire
50
51 221 (two adult admissions units and one children’s assessment unit). The acute units routinely assess and
52
53 222 treat cardio-respiratory admissions due to breathlessness in adults and children.
54
55
56
57
58
59
60

1
2
3 223 Participants with self-reported acute breathlessness, either requiring admission or a change in baseline
4
5 224 treatment, will be screened for the study. Informed consent will be obtained in all participants
6
7 225 following a clinical review by a senior decision maker within 24 hours of acute admission (**Figure 3**).
8
9
10 226
11
12 227

13 228 **2.2. Objectives**

14 229 **2.2.1. Primary objective**

- 15
16 230
17 231
- 18 231 • To evaluate the sensitivity, specificity, positive and negative predictive value of exhaled
19
20 232 breath VOC biomarkers to differentiate acute breathlessness in cardio-respiratory patients.
21

22 233 **2.2.2. Secondary objectives:**

- 23 234
24 235
- 25 235 • To replicate selected breath VOC biomarkers identified in acute breathlessness.
26
27 236 • To discover and replicate breath VOC biomarkers that differentiate the common cardio-
28
29 237 respiratory conditions that cause acute breathlessness, specifically (i) acute heart failure, (ii)
30
31 238 community acquired pneumonia, (iii) adult exacerbations of asthma and COPD and age-
32
33 239 matched adults that do not have cardio-respiratory disease or breathlessness.
34
35 240 • To quantify the level of clinical uncertainty in the primary diagnosis using a 100 mm visual
36
37 241 analogue scale and independent clinical adjudication of case notes blinded to the following
38
39 242 blood biomarkers (i) CRP (ii) BNP (iii) Troponin-I (iv) blood eosinophils, but not clinical
40
41 243 history and acute presentation nor chest x-ray imaging. Potential discriminatory breath VOC
42
43 244 biomarkers will be adjusted for clinical uncertainty in statistical models.
44
45 245 • To identify and replicate exhaled breath VOC biomarkers in school age children treated in
46
47 246 hospital for severe asthma attacks and compare these to age-matched healthy controls.
48
49

50 247 **2.2.3. Exploratory end points (where applicable):**

- 51 248
52
53 249
- 54 249 • To evaluate the dynamic profile of selected breath VOC between the acute state and the
55
56 250 recovery state post exacerbation.
57
58
59
60

- 1
2
3 251 • To evaluate the relationship between exhaled VOC biomarkers and clinical outcomes
4
5 252 including (i) hospital readmission at 30 and 60 days post event (ii) all-cause mortality over a 2
6
7 253 year period post admission.
8
9 254 • To evaluate the relationship between breath VOC biomarkers and functional measures
10
11 255 e.g. physical performance and activity
12
13 256 • To explore potential breath VOC biomarkers of multifactorial acute breathlessness
14
15 257 • To evaluate the relationship between diet, lifestyle and environment upon breath VOC
16
17 biomarkers
18
19
20
21
22

2.3. Sample size estimation

23
24 261
25
26 262 Preliminary data was utilised to conduct sample size estimates from a cohort of acutely breathless
27
28 263 patients admitted to acute admissions units over a 6 month period (February 2017 to August 2017).
29
30 264 Hundred and twelve adult participants (asthma 46, community acquired pneumonia 26, COPD 22)
31
32 265 and eighteen healthy controls were utilised for the analysis.
33
34

35 266 A panel of ten pre-specified aldehydes, based on literature search [31, 39, 40], were extracted from
36
37 267 breath using gas chromatography-mass spectrometry (GC-MS). The aldehydes were normalised to a
38
39 268 common internal standard and were not background-subtracted.
40
41

42 269 A closed formula from Hsieh *et al*[41], relating sample size to observable effect size, was used to
43
44 270 calculate sample size from logistic regression models of the ten aldehydes with acute breathlessness
45
46 271 as the outcome measure. The sample size estimates are also relevant to acute class comparisons vs.
47
48 272 the sum of other acute classes.
49
50

51 273 Based upon the sample size estimates we would have an 80% power, with a type 1 error rate of 5%, to
52
53 274 detect an odds ratio of association of 1.2 between two disease classes with 55 patients per class. Given
54
55 275 the fact that study seeks to discover and replicate breath VOC amongst five adult disease classes
56
57 276 (community acquired pneumonia, heart failure, COPD, asthma and healthy aged matched subjects) we
58
59 277 would require 110 adult patients per class – 550 patients across the program to achieve these aims.
60

1
2
3 278 The closed formulae by Tihaki *et al*,[42] were also utilised to understand the discriminatory power
4
5 279 that the samples sizes above would provide with respect to biomarker sensitivity and specificity; The
6
7 280 following assumptions were made:

- 8
9
10 281 • That a sensitivity of 80 % with a precision of 5% would provide a useful biomarker capable
11
12 282 of ‘ruling out’ an acute class. The same target was applied to specificity.
- 13
14 283 • We assume a prevalence of acute breathlessness of 80% as the recruitment campaign uses
15
16 284 acute breathlessness as the initial stratification tool for recruitment and 1:5 patients recruited
17
18 285 will be non-breathless healthy controls
- 19
20
21 286 • We aim to balance group sizes across classes equally

22
23 287 For a type 1 error rate of 0.05 and a 95% confidence interval

24
25 288 $N_{\text{sensitivity}} = 307$

26
27 289 $N_{\text{specificity}} = 1,230$

28
29 290

30 291 For a type 1 error rate of 0.05 and a 90% confidence interval

31
32 292 $N_{\text{sensitivity}} = 218$

33
34 293 $N_{\text{specificity}} = 871$

35
36 294

37 295 For a type 1 error rate of 0.05 and an 85% confidence interval

38
39 296 $N_{\text{sensitivity}} = 166$

40
41 297 $N_{\text{specificity}} = 664$

42
43 298

44 299 For a type 1 error rate of 0.05 and an 80% confidence interval

45
46 300 $N_{\text{sensitivity}} = 131$

47
48 301 $N_{\text{specificity}} = 524$

49
50 302

51 303 Therefore, we are powered to identify sensitive biomarkers ($\geq 80\%$) of acute breathlessness with a
52
53 304 maximum marginal error in the estimate for sensitivity not exceeding 5% with 95% confidence.

54
55 305 Similarly, we are powered to identify specific biomarkers ($\geq 80\%$) of acute breathlessness with a
56
57 306 maximum marginal error in the estimate for specificity not exceeding 5% with 80% confidence.

1
2
3 307 For the primary analysis the outcome will be treated as a nominal variable with levels (i) acute heart
4
5 308 failure (ii) community acquired pneumonia (iii) adult exacerbations of asthma and COPD and (iv) acute
6
7 309 exacerbations in school age children treated in hospital for severe asthma attacks.
8
9

10 310 The relationship between the primary outcome and the exhaled breath VOC biomarkers will be
11
12 311 modelled using multinomial logistic regression. In addition to metabolomics markers the following
13
14 312 independent variables will be included in the model: clinical uncertainty score on a 100 mm VAS
15
16 313 scale, age, and a validated co morbidity score (the Charlson comorbidity score)[43, 44].
17
18 314

19
20 315 Receiver operator analyses will be utilised to generate ROC curves for individual and multiple panels
21
22 316 VOC predictors in the primary analysis.
23
24 317

25
26
27 318 To understand the dynamic profile of breath biomarkers during (i) the acute state and (ii) in the
28
29 319 chronic state up to 6 months post exacerbation, a repeated measures model with a random intercept
30
31 320 and random effect for time will be fitted, the random effects will be fitted for each patient. For the
32
33 321 repeated measures mixed model an unstructured covariance will be assumed. To evaluate the
34
35 322 relationship between breath biomarkers and hospital readmission at 30 and 60 days Cox proportional
36
37 323 hazards and frailty models will be utilised [45]. Analysis of Multivariate Survival Data, [CITE]
38
39 324 competing risk models and joint models will be fitted [46]. Relationship between death and breath
40
41 325 biomarkers will be evaluated using a logistic regression model. Changes in outcome measures will be
42
43 326 measured appropriately for each variable (e.g. paired t-test, Mann-Whitney, repeated measures
44
45 327 analysis). Tables of descriptive statistics will be compiled for all key variables
46
47
48

49 328 All analysis will be performed using R 3.5.0 <https://www.r-project.org/>.

50
51 329

52 330 **2.4. Discovery and Replication studies**

53
54
55 331 Specific indicator conditions have been selected for targeted recruitment according to their high
56
57 332 prevalence and unmet need, their high morbidity and mortality and the need to develop better
58
59 333 diagnostic and prognostic algorithms in acute care pathways.
60

1
2
3 334 The indicator diagnoses of interest are (i) exacerbations of adult asthma and COPD, (ii) community
4
5 335 acquired pneumonia (iii) acute heart Failure and (iv) exacerbation in school age children treated in
6
7 336 hospital for severe asthma attacks.
8
9

10 337 Patient level clinico-pathological and outcome data (spanning the entire acute pathway) will be
11
12 338 collected in parallel to breath sampling. In addition, breath samples will be acquired in the stable state
13
14 339 post exacerbation (**Figure 3**).
15

16 340
17
18 341 Age matched healthy volunteers will be recruited where possible at separate visits. For the purposes of
19
20 342 this study, healthy volunteers will be defined as participants who have no prior history of asthma,
21
22 343 COPD, heart failure and have not been admitted to hospital with community acquired pneumonia
23
24 344 within 6 weeks of the baseline study visit. For acute admission the study team will approach the
25
26 345 spouse, parent or sibling of the index case and seek informed consent for study assessments. All
27
28 346 healthy subjects will undergo two assessments separated by a duration of 8-16 weeks to match the
29
30 347 acute and recovery time points elapsed in their index case/partner/spouse/sibling/child. Additional
31
32 348 healthy volunteers will be identified from local recruitment databases and via advertising
33
34
35
36

37 350 **2.4.1. Discovery Phase (Project months 1-24):**

38
39 351 The aim of the discovery phase is to discover putative discriminatory breath VOC, using both offline
40
41 352 and online technologies.
42

43
44 353 Pre-planned recruitment of acutely breathless patients will be enriched into the following disease
45
46 354 strata following senior clinical decision maker assessment and within 24 hours of acute admission.
47

48 355 Acute adult heart failure (n=55), adult community acquired pneumonia (n=55), adult exacerbations of
49
50 356 asthma (n=55) and COPD (n=55), acute severe asthma attacks in school age children (n=50).
51

52 357 Additional age matched healthy volunteers (n=55 adults and 50 children) will be identified as a non-
53
54 358 disease reference group (**Table 1**).
55

56 359

57
58 360
59
60

361 **2.4.2. Replication Phase (years 3-4)**

362 The aim of the replication phase is to replicate putative discriminatory breath VOC /VOC signatures
363 identified in the discovery phase.

364 Similar to the discovery phase, recruitment of acutely breathless patients will be enriched into the
365 following disease strata following senior clinical decision maker assessment and within 24 hours of
366 acute admission.

367 Acute adult heart failure (n=55), adult community acquired pneumonia (n=55), adult exacerbations of
368 asthma (n=55) and COPD (n=55), acute severe asthma attacks in school age children (n=25) (**Table**
369 **1**).

370 Additional age matched healthy volunteers (n=55 adults and 25 children) will be identified as a non-
371 disease reference group.

372 **Total combined sample size of the discovery and replication phases = 700 participants**

Disease Category	Discovery	Replication
Acute Adult Asthma	55	55
Acute COPD	55	55
Acute Heart Failure	55	55
Community Acquired Pneumonia	55	55
Adult healthy volunteers	55	55
Acute paediatrics Asthma	50	25
Paediatrics healthy volunteers	50	25
Total sample	375	325

373 **Table (1): Table summarising recruitment targets for both adult and paediatric groups.**

374 **2.5. Schedule of assessments**

375 A schedule of acute assessments is outlined below and aligns to the movement of acute patients
376 through the clinical care pathway and the overall aim of developing a complete phenotypic picture of
377 acutely breathless patients.

378

379 **2.5.1. Defining acute breathlessness**

380 At presentation (within 24 hours of admission) to one of three acute admissions units potentially
381 eligible patients will be identified following confirmation of acute breathlessness, identified as (i)
382 patient defined acute breathlessness and/or (ii) 1 unit increase above patient reported baseline in the
383 extended medical research council (eMRC) dyspnoea score [47, 48] and at least one of the indicator

384 diagnoses identified as the primary clinical diagnosis by a senior clinical decision maker. eMRC will
385 be completed by all patients and healthy volunteers at each research visit.

386 **2.5.2 Informed consent**

387 Patients meeting the pre-specified definition of acute breathlessness will be approached for informed
388 consent in to the breath VOC biomarker study. Only patients that are eligible to give full written
389 informed consent will be recruited.

390 **2.5.3 Collection of blood based pathology markers**

391 Collection of the blood biomarkers CRP, BNP, Troponin-I and blood eosinophil count will be
392 performed both acutely and following recovery, when not taken as part of clinical care pathway.
393 These are currently used in profiling acutely breathless patients in clinical practice (**Table2**).

394

Test	ANALYSER/METHOD	LOWER LIMIT OF DETECTION	UPPER LIMIT OF DETECTION
C-Reactive protein (CRP)	Siemens Advia Chemistry XPT, PEG enhanced immunoturbidimetric. Siemens Advia 1800, PEG enhanced immunoturbidimetric	5 mg/L	Diluted to result
B-type natriuretic peptide (BNP)	Siemens Advia Centaur XPT, two-site sandwich immunoassay using direct chemiluminescent technology	2.0 pg/mL	1445 pg/mL
Troponin-I	Abbott Architect i2000SR, three-site sandwich immunoassay using direct chemiluminescent technology (CMIA).	5.0 ng/L	50,000 ng/L

395 **Table (2):** Type of analyser and methodology used for blood biomarker calculation. The table
396 outlines analyser make, methodology, upper and lower limits of detection as per the University
397 Hospitals of Leicester NHS Foundation trust laboratory guidelines.

398

399 **2.5.4 Breath VOC sampling**

400 Offline breath sampling using GC-MS and comprehensive two-dimensional gas chromatography-
401 mass spectrometry coupled with a standardised and CE marked breath sampler-ReCIVA[®][49], will be
402 performed. Gas chromatography is considered a gold standard technique in detecting volatile organic
403 compounds and as such its sampling will be prioritised. Additionally the following online
404 technologies, proton transfer mass spectroscopy (PTR-MS), gas chromatography - ion mobility

405 spectroscopy (GC-IMS) and atmospheric pressure chemical ionisation- mass spectrometry (APCI-
 406 MS) will be evaluated according to the sampling strategy outlined in section 3, **Figure 3 and Table 3.**

	COPD		Asthma		Pneumonia		Heart Failure		Healthy		Paediatrics	
	1	2	1	2	1	2	1	2	1	2	1	2
Time point												
Written informed consent	x		x		x		x		x		x	
Volatile organic compound (VOC) sampling												
ReCIVA- gas chromatography and mass spectrometry (GC-MS)	x	x	x	x	x	x	x	x	x	x	x	x
ReCIVA- comprehensive two-dimensional gas chromatography (GCxGC-MS)	x	x	x	x	x	x	x	x	x	x	x	x
Atmospheric pressure chemical ionisation- mass spectrometry (APCI-MS)	x	x	x	x	x	x	x	x	x	x	x	x
Proton transfer reaction mass spectrometry (PTRMS)		x		x		x		x	x	x	x	x
Gas chromatography - ion mobility spectrometry (GC-IMS)	x		x		x		x				x	x
Pathology blood tests												
Full blood count (including differential cell count)	x	x	x	x	x	x	x	x	x	x	x	x
Brain natriuretic peptide (BNP) [pg/mL]	x	x	x		x		x	x	x			
Troponin-I [ng/L]	x		x		x		x		x			
C-Reactive protein (CRP) [mg/L]	x	x	x	x	x	x	x	x	x	x	x	x
Lung function tests												
Hand held forced oscillation technique (FOT)	x	x	x	x	x	x	x	x	x	x		
Fractional exhaled nitric oxide (FeNO) - Flow rate 50 [ml/s]			x	x							x	x
Spontaneous sputum sample	x	x	x	x	x	x	x	x	x	x		
Bio-banking (urine, serum, plasma. sputum supernatants & plugs)	x	x	x	x	x	x	x	x	x	x	x	x
Transthoracic echocardiography	x						x					

407 **Table (3):** Summary of baseline and follow up assessments. The table summarises key assessments
 408 carried out at different time points during the study. The participants may undertake any combination
 409 of the investigations listed at any of these time points.

410

411

2.5.5. *Collection of additional samples for future biomarker campaigns*

Collection of additional biomarkers for future biomarker discovery campaigns including (i) a urine sample, (ii) blood samples, up to 85mls for DNA, RNA, plasma and serum and peripheral blood cell flow cytometry in selected subjects and (iii) spontaneous or induced sputum samples (plugs and supernatants) will be carried out (**Table 3**).

All samples will be collected at time point 1 and at 2 (**Figure 3**). The additional samples will be used for future omics analyses, these may include detailed analysis of the metagenome in sputum and proteomics applied to urine and serum samples.

2.5.6. *Physiological characterisation*

Physiological measures of lung function will be performed in acutely ill participants and at recovery including, (i) Hand held forced oscillation technique (FOT): an easily accessible measure of lung function. Patients favour this to spirometry as it is effort independent, unlike spirometry and requires less than a minute of quiet tidal breathing to obtain triplicate high quality measurements [50], This will be completed using Tremoflo®, Thorasys Thoracic Medical Systems Inc. (ii) Fractional exhaled nitric oxide (FeNO): A measure of airway inflammation in asthmatic patients[51, 52]. This instrument used for this will be NIOX VERO®, registered trademark of Circassia AB (PP-VERO-UK-0022-v1.0) (iii) Echocardiography: Two dimensional transthoracic echocardiography will be performed in heart failure and COPD patients using an iE 33 system with S5-1 transducer (frequency transmitted 1.7 MHz, received 3.4 MHz; Philips Medical Systems, Best, The Netherlands). Standard techniques as per American Society of Echocardiography guidelines (ASE)[53] will be used to acquire 2D, colour and Doppler images in conventional parasternal long-axis, short-axis and apical 4-, 2- and 3-chamber views. Left ventricular ejection fraction (LVEF)

1
2
3 437 will be calculated using the biplane method of discs formula (Simpson's rule) to derive left
4
5 438 ventricular volume indices.
6
7
8

9 439 All participants are encouraged to report any testing related discomfort or concerns to the research
10
11 440 team to terminate the sampling process.
12

13 441 **2.5.7 Recovery follow up**

14
15 442 • Patient recovery will be defined as:

16
17 443 (i) Patient reported recovery from the acute exacerbation spell and back to their
18
19 444 baseline extended MRC score or clinician defined recovery from the acute
20
21 445 exacerbation spell
22

23
24 446 and

25
26 447 (ii) At least 6 weeks post exacerbation event (up to 6 months).
27

28 448

29
30 449 Patients that re admit to hospital between visits 1 and 2, can have additional visit 1 assessments. Visit
31
32 450 2 will be taken as recovery following the subsequent admission. If a patient is admitted to hospital
33
34 451 after visit 2 then they will be eligible to be recruited as a new study participant.
35

36 452

37
38 453 The schedule of assessments at the recovery visit is outlined (**Table 3**).
39

40 454

41 455 **2.6. Clinical Adjudication:**

42
43 456

44
45 457 In an effort to reduce data variability and minimise bias, an independent panel consisting of two senior
46
47 458 acute clinicians (SS & NG) will review all pertinent clinical and diagnostic source documentation,
48
49 459 whilst blinded to admission blood biomarkers and clinical diagnosis.
50

51
52 460 All acutely breathless adult patients' notes will be sequentially divided for adjudication. The panel will
53
54 461 independently determine the primary diagnosis of highest probability from a list of the four potential
55
56 462 acute indicator diagnoses in section 2.4 and mark their level of clinical certainty on a 100 mm visual
57
58 463 analogue scale (VAS scale). The panel members will be able to review imaging, electrocardiograms
59
60

1
2
3 464 (ECGs), and other relevant information but not admission blood based pathology tests.
4

5 465 In a subset of patients adjudication will be validated by separate panel member to ensure between
6
7 466 observer agreement using Bland-Altman analysis and inter-rater agreement of the primary diagnosis
8
9 467 using Kohen's kappa via repeated evaluation of a subset of cases (see statistical methods section 2.3).
10

11 468

14 469 **2.7. Clinical Informatics:**

15 470

17 471 Clinical data collection will be undertaken using a securely-hosted bespoke database system (ADD)
18
19 472 developed within the NIHR Leicester Biomedical Research Centre – Respiratory (BRC). The system
20
21 473 links acute admission episodes to hospital pathology records; historical respiratory physiology tests;
22
23 474 and demographic information. The system provides functionality to validate data entry; manually
24
25 475 verify records and highlight incomplete records. A custom VOC 'module' has been created to
26
27 476 support data collection within the study visits (1 and 2), and standardise diagnoses and medications
28
29 477 through the use of clinical ontologies as well as linking hospital records/tests to patient visits.
30

31 478

33 479 Non-ADD based clinical data (e.g. hospital admissions, re-admissions and mortality) will be extracted
34
35 480 from the hospital data warehouse using identifiable patient identifiers, and subsequently
36
37 481 pseudonymised prior to integration.
38

39 482

42 483 An informatics pipeline will be created to facilitate the transfer of chemo-metric data from remote
43
44 484 computers to the data repository. This will include tools to (i) enforce the correct labelling of data sets
45
46 485 (e.g. study number, visit, type/source of sample) prior to automated validated transfer to the
47
48 486 repository; (ii) record information about the sample process; (iii) search and extract data sets from the
49
50 487 repository for subsequent analysis. Prior to analysis, clinical and chemo-metric data will be integrated
51
52 488 using the study number and any potentially identifiable information will be removed.
53

54 489

55 490

56 491

3. Breath profiling

The technologies utilised in the VOC study during discovery and replication phases are:

Offline technologies

- ReCIVA+ GC-MS
- ReCIVA + GC x GC-MS
-

Online technologies

- GC-IMS
- PTR-MS
- APCI-MS

Offline technologies will underpin the discovery analyses owing to their ability to identify chemical identity and their recognition as the analytical gold standard in exhaled breath VOC analysis [54].

In contrast online technologies will be utilised for VOC biomarker replication and at the recovery visits owing to their portability and potential for future point of care testing. **(Figure 4)**.

A brief description of the core VOC platforms is provided below

A CE marked breath sampling device (ReCIVA®) developed by Owlstone Medical, will be used to sample breath onto two adsorbent Tenax tubes. Participants will be asked to breathe through the ReCIVA face mask for a maximum of 300 seconds, aiming for collection of $\geq 80\%$ of the target sample volume of 1 litre, after which the Tenax tubes will be transferred to the laboratory for analysis. This effectively allows de-coupling of the breath sampling from the breath sensor and analysis platforms in selected patients that are not able to mobilise to a real time breath sampling device. The Owlstone ReCIVA sampler will be utilised in breath collection for offline technologies namely GC-MS and GCxGC-MS. The ReCIVA sampler is capable of entraining oxygen and is therefore suitable for patients with mild respiratory failure requiring low flow rates of oxygen to maintain target oxygen saturations [49].

1
2
3 519 **3.1. Gas chromatography and mass spectroscopy (GC-MS):** is a commonly applied methodology
4
5 520 used to accurately measure trace gases in complex mixtures such as exhaled air [54]. Pre-
6
7 521 concentrating breath volatiles by various means and subsequent analysis constitute a reliable and
8
9 522 sensitive method for VOC analysis [55]. Despite its high sensitivity, it is however, a time consuming
10
11 523 technique and carries a risk of contamination at the pre-concentration step. It is also not suitable for
12
13 524 online and multiple measurements limiting its use as a point-of-care testing technology for VOC [56].
14
15
16 525 The instrument used will be an Agilent 7890A gas chromatogram with a 5977a quadrupole mass
17
18 526 spectrometer (Agilent Technologies Ltd, Stockport, UK), interfaced with a Markes Unity 2 thermal
19
20 527 desorptionunit (Markes International Ltd, Llantrisant, UK).

21
22
23 528 **3.2. Comprehensive two-dimensional gas chromatography-mass spectrometry (GCxGC-MS):** an
24
25 529 advanced analytical technique for the analysis of complex organic matrices; its main advantage is the
26
27 530 unparalleled separation power it affords over conventional one-dimensional chromatographic
28
29 531 techniques [57]. Previous research, albeit sparse, has demonstrated the potential of GCxGC-MS for
30
31 532 breath analysis with the number of VOC detected exceeding those detected by conventional GC-MS
32
33 533 [58, 59]. GCxGC-MS of breath metabolites has been used for the identification of biomarkers related
34
35 534 to glucose metabolism [60, 61], tuberculosis [62] and radiation response [63]. This has generated
36
37 535 interest within the breath research community, however, such studies were conducted on a small scale
38
39 536 (<50 patients) and involved the use of expensive detectors and modulators. Method development and
40
41 537 analysis of the data-rich GCxGC chromatograms, however, can be time-consuming and require
42
43 538 specialist knowledge.

44
45
46
47 539 The instrument used will be an Agilent 7890A gas chromatogram, fitted with a G3486A CFT flow
48
49 540 modulator and a three-way splitter plate coupled to a flame ionisation detector and a HES 5977B
50
51 541 quadrupole mass spectrometer (Agilent Technologies Ltd, Stockport, UK), interfaced with a Markes
52
53 542 TD-100xr thermal desorption autosampler (Markes International Ltd, Llantrisant, UK).

54
55
56 543 **3.3. Proton-transfer-reaction mass spectrometry (PTRMS):** a real time technique, capable of
57
58 544 simultaneously measuring the evolution of multiple gas metabolites from a single breath. It has been
59
60

1
2
3 545 used for the identification of potential useful VOC biomarkers for diagnosis of a variety of diseases:
4
5 546 including various cancers[64-66], liver disease[67, 68] and respiratory disease[69]. It has several
6
7 547 advantages in clinical settings, such as the speed of sampling, the instant result achieved and the lack
8
9 548 of need for sample storage or shipping. However, owing to the lack of pre-concentration or
10
11 549 chromatographic separation, sensitivity and definitive compound identification can be somewhat
12
13
14 550 limited when compared to GC-MS.

15
16 551 Two breath sampling devices will be used. The first device is a Loccioni SOFIA GSI-S; the subject is
17
18 552 required to exhale a single breath, five times (three if providing five samples proves too difficult) into
19
20 553 a sterile mouthpiece connected to an electrostatic bacterial/viral filter whilst wearing a nose clip (all
21
22 554 CE marked). Flow from the mouthpiece passes into a gas sampling interface capnograph (Loccioni
23
24 555 GSI-S – CE marked) and real-time user feedback of flow is provided on screen, allowing the
25
26 556 regulation of the breath sampling rate. The gas sampling interface acts to simultaneously trigger the
27
28 557 acquisition of the PTR-ToF-MS data and the exhaled breath travels through the capnograph down a
29
30 558 heated sample line into the ion source of the PTR-ToF-MS

31
32
33
34 559 The second breath sampling device is a ReCIVA breath sampler (Owlstone) with one of the adsorbent
35
36 560 Tenax tubes replaced with an outlet tube adapted for online sampling. The exhaled breath is
37
38 561 transferred to the PTR-ToF-MS via a heated transfer line connected to the outlet tube, continuously
39
40 562 drawn at a constant flow rate by the PTR-ToF-MS. The online adaptation of the consumable
41
42 563 adsorbent tube does not affect the CE mark of the ReCIVA sampling device.

43
44
45 564 Once the breath sample reaches the PTR-ToF-MS, via either breath sampler, the breath mixes with
46
47 565 protonated water (H_3O^+) inducing proton transfer to the target volatile organic compounds (VOCs)
48
49 566 present, resulting in their ionisation. Sample ions are then guided into the time of flight mass
50
51 567 spectrometer and mass spectra, showing the abundance and mass of the VOCs present, are collected
52
53 568 throughout the exhalation. Following sampling, mouthpieces, filters and nose clips are disposed of
54
55 569 and all patient contacted surfaces wiped down with antiseptic cleaning wipes in preparation for the
56
57
58 570 next patient.

1
2
3 571 The instrument used will be a Kore Series II high performance proton transfer reaction-time of flight-
4
5 572 mass spectrometer (Kore Technology Ltd, Cambridge, UK).
6
7

8 573
9

10 574 **3.4. Gas chromatography- ion mobility spectrometry (GC-IMS) (B&S Analytik):** Allows the
11
12 575 detection of volatile organic compounds down to ultra-trace level (ng/L - to pg/L - range). For years,
13
14 576 IMS has been used to discover potential discriminatory breath VOC in lung cancer [70, 71],
15
16 577 COPD[72, 73] and asthma[73]. Sampling takes place using a SpiroScout spirometer. The patients
17
18 578 exhale through a disposable mouth piece connected to a Teflon tube. A piezoelectric pressure sensor
19
20 579 is used to monitor the breathing profile, this opens the sampling valve at the appropriate point in the
21
22 580 breath profile to collect end-tidal breath in a sample loop of 10 mL volume. After filling this loop, the
23
24 581 collected sample air is then transferred to a multicapillary column for a chromatographic separation,
25
26 582 which is achieved in 12 min. The separated molecules are then transferred into the IMS, ionised and
27
28 583 then separated according to their mobility in a weak electric field.
29
30 584 The technology's multiple advantages of ultra-sensitivity, portability, online sampling and short
31
32 585 analysis time (typical analysis time of 10 minutes) with real time detection, brings a promise to
33
34 586 provide immediate and potentially reliable results for point of care breath diagnostics. Another
35
36 587 concept with IMS devices is that once the required breath signatures have been discovered using GC-
37
38 588 MS, IMS offers the potential to be 'tuned' for selective detection of VOC.
39
40
41
42

43 589 The instrument used will be a BioScout a multi-capillary column gas chromatogram-ion mobility
44
45 590 spectrometer, with a ⁶³Ni ion source, interfaced with a SpiroScout breath sampler (BS Analytik,
46
47 591 Dortmund, Germany).
48
49

50 592
51

52
53 593 **3.5. Atmospheric pressure chemical ionisation- mass spectrometry (APCI-MS) semiportable**
54
55 594 **compact version (Advion):** is one of less sensitive but more affordable versions of mass spectrometers
56
57 595 released to the commercial market in recent years. The device uses APCI to produce ions. Although
58
59 596 the most common use of APCI-MS systems is the detection in liquid chromatography applications,
60

1
2
3 597 the technique has proven to be a valuable tool for direct measurement of VOC in air[74, 75] food[76,
4
5 598 77] and breath[78, 79]. Recently, the technique has shown potential for online, real time profiling of
6
7 599 pseudo-metabolites in exhaled breath [80] with sensitivity comparable with other techniques. By
8
9 600 combining miniaturised MS technology with APCI techniques, adequate quality of on-site, real time
10
11 601 measurements with minimal or no sample preparation requirement can be provided. This is a desirable
12
13 602 outcome as it overcomes main limitation of using standard breath analysis method in clinical setting,
14
15 603 which is a need for breath sample collection followed by desorption and time-consuming laboratory
16
17
18 604
19
20 605
21 637 There remains an overall lack of standardisation and rigour across these technologies which hindered
22
23 638 previous advancements in breath discovery; something we intend to minimize.

24
25 639 The instrument used will be an Advion Compact Mass Spectrometer Express, with atmospheric
26
27 640 pressure chemical ionisation, interfaced with a heated breath sampling line (Advion, New York,
28
29 641 USA).

30
31
32
33 642

34 35 643 **4. Chemometric processing and data analysis:**

36
37
38 644 GC-MS breath data will be aligned, deconvoluted and the features for each participant will be
39
40 645 extracted. The extracted features will be grouped and classified by retention index and mass spectrum.
41
42 646 The registered and aligned data will be linked to participant meta-data to generate a breath matrix.
43
44 647 Data handling and analysis will be performed by a senior statistician.

45
46
47 648 The breath matrix is a $n \times p$ matrix where n is the number of subjects and p is the number of VOC. The
48
49 649 breath matrix is high dimensional with $p \gg n$ and many potentially correlated VOC. In view of this, we
50
51 650 will employ sparse partial least squares discriminant analysis (sPLS-DA)[81] to investigate which of
52
53 651 the VOC can identify breathlessness. We will also investigate which of the VOC can discriminate
54
55 652 between the different disease states including acute exacerbations of asthma and COPD and

1
2
3 653 Pneumonia. In addition to the supervised methods, unsupervised methods will be explored,
4
5 654 specifically sparse principle component analysis (sPCA)[82].
6
7
8 655 Extracted VOC will also be investigated. Relationships between VOC and patient reported acute
9
10 656 breathlessness will be analysed using logistic regression model. VOC associated with patient
11
12 657 associated acute breathlessness will be incorporated into multinomial logistic regression models in
13
14 658 conjunction with CRP, BNP, blood eosinophils and Troponin-I, pathology biomarkers currently in use
15
16 659 for diagnosing undifferentiated breathlessness. In addition to the conventional binary and multinomial
17
18 660 logistic regression models, regression models [83].
19
20
21

22 661 **5. Ethics and dissemination:**

23
24
25 662 The study has obtained full ethical approval from the London South East Research ethics Committee,
26
27 663 REC reference 16/LO/1747. IRAS project ID 198921. Publications will be prepared according to the
28
29 664 MRC-EMBER consortium agreement and the University of Leicester publications policy. All intended
30
31 665 publications will be submitted to the EMBER executive board for review and comments within 60 days
32
33 666 of journal submission. Authorship will be according to contribution and internationally recognised
34
35 667 guidance on journal authorship.
36
37
38

39 668 **6. Study dates:** 01/2/2017 – 30/10/2020

40 669 **7. Authors' contributions:**

41
42
43 670
44
45
46 671 S.S, C.E.B, N.Gr, P.Th and P.Mo conceived the study, obtained funding, wrote the study protocol,
47
48 672 obtained ethical and MHRA approvals for the study and coordinated the deployment of analytical
49
50 673 testing methods for breath analysis. W.I took the lead in writing the manuscript with support from
51
52 674 S.S. Planning and recruitment of adult participants was carried out by W.I, S.Jo, B.Pa, A.Aw, R.Ph,
53
54 675 G.Fo, A.Yo, R. J. R and C.Wh. Paediatrics study design was conceived by E.Ga and C.Be and
55
56 676 participants recruited by T. Mc and C. Fo. Analytical chemistry team formed of M.Wi, R.Co, D.Sa,
57
58 677 D.Ru and L.Br expertly handled all the breath samples and planned an analysis structure. M.Ri, a
59
60

1
2
3 678 senior statistician, constructed a statistics and data analysis plan in conjunction with SS.
4
5 679 Bioinformatics pipeline and electronic CRFs developed by R.Fr and B.Zh. All authors, including
6
7 680 R.Pe, H. Bh, B.Ha, A. Si, K. Ry, H. Pa, T. Su, L. L Ng, T. Co contributed to the study design and
8
9 681 study protocol.
10

11
12
13 682 **8. Protocol version:** Version 4, 1st April 2018
14

15
16 683

17
18 684 **9. Public and patient involvement:**

19
20 685 A series of consultations have taken place with our patient involvement team within the NIHR Biomedical
21
22 686 Research Centre (Respiratory Theme) and across the wider BRC PPI group. Representations from the
23
24 687 paediatrics team were also present. This group was sent copies of the participant documentation for review
25
26 688 and discussion. Various revisions have been made following on from these discussions.
27
28

29
30 689 **10. Funding statement:**

31
32 690 Research funding was provided by the MRC and EPSRC Stratified Medicine Grant for Molecular
33
34 691 Pathology Nodes (Grant No. MR/N005880/1). The research was supported by the National Institute
35
36 692 for Health Research (NIHR) Leicester Biomedical Research Centre and NIHR, Leicester Clinical
37
38 693 Research Facility and the Midlands Asthma and Allergy Research Association (MAARA) to whom
39
40 694 we are extremely grateful. The authors would like to acknowledge the invaluable efforts of the
41
42 695 research nurses responsible for the in-clinic sample collection as well as the input from the wider
43
44 696 EMBER consortium (Members list can be found at: <https://ember.le.ac.uk/web>). The views expressed
45
46 697 are those of the author(s) and not necessarily those of the NHS and NIHR or the Department of Health
47
48

49
50 698

51
52 699 **11. Competing interests:** SS has performed advisory services for Owlstone Medical.
53

54
55 700

56
57 701

58
59 702
60

1
2
3 **703 Figure legends:**
4

5 **704 Figure (1)** Relationship between lung proximity and degree of invasiveness of different lung
6
7 **705** matrices. The figure plots the level of invasiveness of various lung matrices in relation to their
8
9 **706** proximity to the lung. Given their pathological relevance, the degree of invasiveness of
10
11 **707** bronchoalveolar lavage (BAL) and lung biopsy makes them less favourable in diagnosing respiratory
12
13 **708** diseases.
14

15
16 **709 Figure (2)** Multi-instrument use in breath sampling. Figure illustrates the various combinations of
17
18 offline and online devices used in breath sampling and the relevant pros and cons. Offline and online
19
20 technologies are used for the discovery and validation phases of the study respectively.
21
22

23
24 **712 Figure (3)** Study flow chart. Figure outlines the patient journey from admission through to discharge
25
26 **713** and follow up. Participants with self-reported acute breathlessness presenting to University Hospitals
27
28 **714** of Leicester are recruited within 24 hours following a senior decision maker review. Breath sampling
29
30 **715** is carried out on the first visit, at recruitment, and the second visit, up to 6 months after discharge.
31
32 **716** Patients are admitted through the standard operational emergency medical streaming and care
33
34 **717** pathways at the University Hospitals of Leicester NHS Trust. Early outcomes (hospital readmission)
35
36 **718** are measured at 30 and 60 days and late outcomes (mortality) including respiratory and all-cause
37
38 **719** mortality are measured at 2 years. Assessments carried out at each time point are summarised in
39
40 **720 Table 1.**
41
42

43
44 **721 Figure (4)** Multi-instrument use in breath sampling. Operational space of the analytical technologies
45
46 **722** used in EMBER for the analysis of volatile organic compounds (VOC) in exhaled breath, including
47
48 **723** proton transfer reaction-mass spectrometry (PTR-MS), atmospheric pressure chemical ionisation-
49
50 **724** mass spectrometry (CMS), gas chromatography-ion mobility spectrometry (GC-IMS), gas
51
52 **725** chromatography-mass spectrometry (GC-MS) and two-dimensional gas chromatography-mass
53
54 **726** spectrometry (GCxGC). Comparing the typical molar mass range detectable; selectivity in detection
55
56 **727** owing to the type of ionisation involved and the proton affinity of analytes; and the inclusion of a
57
58 **728** chromatographic separation affecting total time of analysis. The online technologies involving
59
60

1
2
3 729 chemical ionisation (PTR-MS, CMS and GC-IMS) can be used in-clinic owing to short analysis
4
5 730 times, but only detect lower molar mass molecules with a proton affinity higher than 697 KJ mol⁻¹.
6
7 731 Offline chromatographic techniques (GC-MS and GCxGC) detect a wider range of compounds
8
9 732 independent of proton affinity; however, the techniques have longer analysis times and involve
10
11 733 sample transportation and storage.
12

13
14 734

15
16 735

17
18 736

19
20 737

21
22 738

23
24 739

25
26 740

27
28 741

29
30 742

31
32 743

33
34 744

35
36 745

37
38 746

39
40 747

41
42 748

43
44 749

45
46 750

47
48 751

49
50 752

51
52 753

53
54 754

55
56 755

57
58
59
60

For peer review only

8. References

1. Stephenson, J.J., et al., *Clinical and economic burden of dyspnea and other COPD symptoms in a managed care setting*. *Int J Chron Obstruct Pulmon Dis*, 2017. **12**: p. 1947-1959.
2. David Ponka, M.K., *Top differential diagnoses in family medicine Dyspnoea Can Fam Physician*. Canadian Family Physician, 2008.
3. Figarska, S.M., H.M. Boezen, and J.M. Vonk, *Dyspnea severity, changes in dyspnea status and mortality in the general population: the Vlagtwedde/Vlaardingen study*. *European Journal of Epidemiology*, 2012. **27**(11): p. 867-876.
4. Parshall, M.B., et al., *An Official American Thoracic Society Statement: Update on the Mechanisms, Assessment, and Management of Dyspnea*. *American Journal of Respiratory and Critical Care Medicine*, 2012. **185**(4): p. 435-452.
5. Lim, W.S., et al., *BTS guidelines for the management of community acquired pneumonia in adults: update 2009*. *Thorax*, 2009. **64 Suppl 3**: p. iii1-55.
6. Leuzzi, G., et al., *C-reactive protein level predicts mortality in COPD: a systematic review and meta-analysis*. *Eur Respir Rev*, 2017. **26**(143).
7. Moe, G.W., *B-type natriuretic peptide in heart failure*. *Curr Opin Cardiol*, 2006. **21**(3): p. 208-14.
8. Ponikowski, P., et al., *2016 ESC Guidelines for the diagnosis and treatment of acute and chronic heart failure: The Task Force for the diagnosis and treatment of acute and chronic heart failure of the European Society of Cardiology (ESC) Developed with the special contribution of the Heart Failure Association (HFA) of the ESC*. *Eur Heart J*, 2016. **37**(27): p. 2129-2200.
9. Alex Sakula, *Charcot leyden crystals and crushmann spirals*. *Thorax*, 1986.
10. Calhoun, W.J., J. Sedgwick, and W.W. Busse, *The role of eosinophils in the pathophysiology of asthma*. *Ann N Y Acad Sci*, 1991. **629**: p. 62-72.
11. Pavord, I.D., *Eosinophilic phenotypes of airway disease*. *Ann Am Thorac Soc*, 2013. **10 Suppl**: p. S143-9.
12. Tashkin, D.P. and M.E. Wechsler, *Role of eosinophils in airway inflammation of chronic obstructive pulmonary disease*. *International Journal of Chronic Obstructive Pulmonary Disease*, 2018. **13**: p. 335-349.
13. Bafadhel, M., et al., *Blood Eosinophils to Direct Corticosteroid Treatment of Exacerbations of Chronic Obstructive Pulmonary Disease: A Randomized Placebo-Controlled Trial*. *American Journal of Respiratory and Critical Care Medicine*, 2012. **186**(1): p. 48-55.
14. Bedossa, P. and K. Patel, *Biopsy and Noninvasive Methods to Assess Progression of Nonalcoholic Fatty Liver Disease*. *Gastroenterology*, 2016. **150**(8): p. 1811-1822 e4.
15. Fens, N., et al., *Exhaled breath analysis by electronic nose in airways disease. Established issues and key questions*. *Clin Exp Allergy*, 2013. **43**(7): p. 705-15.
16. Probert, C.S., et al., *Volatile organic compounds as diagnostic biomarkers in gastrointestinal and liver diseases*. *J Gastrointestin Liver Dis*, 2009. **18**(3): p. 337-43.
17. Schubert, J.K., et al., *Breath analysis in critically ill patients: potential and limitations*. *Expert Rev Mol Diagn*, 2004. **4**(5): p. 619-29.
18. Fowler, S.J., et al., *Surveillance for lower airway pathogens in mechanically ventilated patients by metabolomic analysis of exhaled breath: a case-control study*. *Thorax*, 2015. **70**(4): p. 320-5.
19. Florian Gahleitner, C.G.-H., Caroline S Beardsmore, Hitesh C Pandya, CL Paul Thomas, *Metabolomics pilot study to identify VOC of childhood asthma in exhaled breath*. *Bioanalysis*, 2013.
20. Van Berkel, J.J., et al., *A profile of volatile organic compounds in breath discriminates COPD patients from controls*. *Respir Med*, 2010. **104**(4): p. 557-63.

- 1
2
3 805 21. Basanta, M., et al., *Exhaled volatile organic compounds for phenotyping chronic obstructive*
4 806 *pulmonary disease: a cross-sectional study*. *Respir Res*, 2012. **13**: p. 72.
- 5 807 22. Besa, V., et al., *Exhaled volatile organic compounds discriminate patients with chronic*
6 808 *obstructive pulmonary disease from healthy subjects*. *International Journal of Chronic*
7 809 *Obstructive Pulmonary Disease*, 2015. **10**: p. 399-406.
- 8 810 23. Hauschild, A.C., J.I. Baumbach, and J. Baumbach, *Integrated statistical learning of metabolic*
9 811 *ion mobility spectrometry profiles for pulmonary disease identification*. *Genet Mol Res*, 2012.
10 812 **11**(3): p. 2733-44.
- 11 813 24. Jareno-Esteban, J.J., et al., *Study of 5 Volatile Organic Compounds in Exhaled Breath in*
12 814 *Chronic Obstructive Pulmonary Disease*. *Archivos De Bronconeumologia*, 2017. **53**(5): p. 251-
13 815 256.
- 14 816 25. Paredi, P., et al., *Exhaled ethane, a marker of lipid peroxidation, is elevated in chronic*
15 817 *obstructive pulmonary disease*. *Am J Respir Crit Care Med*, 2000. **162**(2 Pt 1): p. 369-73.
- 16 818 26. Caldeira, M., et al., *Profiling allergic asthma volatile metabolic patterns using a headspace-*
17 819 *solid phase microextraction/gas chromatography based methodology*. *J Chromatogr A*, 2011.
18 820 **1218**(24): p. 3771-80.
- 19 821 27. Caldeira, M., et al., *Allergic asthma exhaled breath metabolome: a challenge for*
20 822 *comprehensive two-dimensional gas chromatography*. *J Chromatogr A*, 2012. **1254**: p. 87-97.
- 21 823 28. Dallinga, J.W., et al., *Volatile organic compounds in exhaled breath as a diagnostic tool for*
22 824 *asthma in children*. *Clin Exp Allergy*, 2010. **40**(1): p. 68-76.
- 23 825 29. Dragonieri, S., et al., *An electronic nose in the discrimination of patients with asthma and*
24 826 *controls*. *J Allergy Clin Immunol*, 2007. **120**(4): p. 856-62.
- 25 827 30. Fens, N., et al., *Exhaled breath profiling enables discrimination of chronic obstructive*
26 828 *pulmonary disease and asthma*. *Am J Respir Crit Care Med*, 2009. **180**(11): p. 1076-82.
- 27 829 31. Ibrahim, B., et al., *Non-invasive phenotyping using exhaled volatile organic compounds in*
28 830 *asthma*. *Thorax*, 2011. **66**(9): p. 804-9.
- 29 831 32. Montuschi, P., et al., *Diagnostic performance of an electronic nose, fractional exhaled nitric*
30 832 *oxide, and lung function testing in asthma*. *Chest*, 2010. **137**(4): p. 790-6.
- 31 833 33. Christiansen, A., et al., *A systematic review of breath analysis and detection of volatile*
32 834 *organic compounds in COPD*. *J Breath Res*, 2016. **10**(3): p. 034002.
- 33 835 34. de Vries, R., et al., *Integration of electronic nose technology with spirometry: validation of a*
34 836 *new approach for exhaled breath analysis*. *J Breath Res*, 2015. **9**(4): p. 046001.
- 35 837 35. Shafiek, H., et al., *Using the Electronic Nose to Identify Airway Infection during COPD*
36 838 *Exacerbations*. *PLoS One*, 2015. **10**(9): p. e0135199.
- 37 839 36. Finamore, P., et al., *Analysis of volatile organic compounds: an innovative approach to heart*
38 840 *failure characterization in older patients*. *J Breath Res*, 2018. **12**(2): p. 026007.
- 39 841 37. *MRC Molecular Pathology, Review 20 Aug.*
- 40 842 38. Medical Research Council, *Molecular Pathology Nodes*. 2014: MRC website.
- 41 843 39. Corradi, M., et al., *Aldehydes in exhaled breath condensate of patients with chronic*
42 844 *obstructive pulmonary disease*. *Am J Respir Crit Care Med*, 2003. **167**(10): p. 1380-6.
- 43 845 40. Larstad, M., et al., *Determination of malondialdehyde in breath condensate by high-*
44 846 *performance liquid chromatography with fluorescence detection*. *J Chromatogr B Analyt*
45 847 *Technol Biomed Life Sci*, 2002. **766**(1): p. 107-14.
- 46 848 41. F. Y. Hsieh, D.A.B., Michael D. Larsen,, *A simple method of sample size calculation for linear*
47 849 *and logistic regression* *Statis Med*, 1998.
- 48 850 42. Hajian-Tilaki, K., *Sample size estimation in diagnostic test studies of biomedical informatics*. *J*
49 851 *Biomed Inform*, 2014. **48**: p. 193-204.
- 50 852 43. Quan, H., et al., *Updating and validating the Charlson comorbidity index and score for risk*
51 853 *adjustment in hospital discharge abstracts using data from 6 countries*. *Am J Epidemiol*,
52 854 2011. **173**(6): p. 676-82.
- 53
54
55
56
57
58
59
60

- 1
2
3 855 44. Roffman, C.E., J. Buchanan, and G.T. Allison, *Charlson Comorbidities Index*. J Physiother, 2016. **62**(3): p. 171.
- 4 856
- 5 857 45. Bjarnason, H. and P. Hougaard, *Fisher information for two gamma frailty bivariate Weibull models*. Lifetime Data Anal, 2000. **6**(1): p. 59-71.
- 6 858
- 7 859 46. Lawrence Gould, A., et al., *Joint modeling of survival and longitudinal non-survival data: current methods and issues. Report of the DIA Bayesian joint modeling working group*. Stat Med, 2015. **34**(14): p. 2181-95.
- 8 860
- 9 861
- 10 862 47. Steer, J., et al., *Dyspnoea severity and pneumonia as predictors of in-hospital mortality and early readmission in acute exacerbations of COPD*. Thorax, 2012. **67**(2): p. 117-21.
- 11 863
- 12 864 48. Tsimogianni, A.M., et al., *Predictors of Outcome After Exacerbation of Chronic Obstructive Pulmonary Disease*. Journal of General Internal Medicine, 2009. **24**(9): p. 1043-1048.
- 13 865
- 14 866 49. Kitchen, S., et al., *LATE-BREAKING ABSTRACT: Breathe free: Open source development of a breath sampler by a consortium of breath researchers*. European Respiratory Journal, 2015. **46**(suppl 59).
- 15 867
- 16 868
- 17 869 50. Oostveen, E., et al., *The forced oscillation technique in clinical practice: methodology, recommendations and future developments*. Eur Respir J, 2003. **22**(6): p. 1026-41.
- 18 870
- 19 871 51. Pendharkar, S. and S. Mehta, *The clinical significance of exhaled nitric oxide in asthma*. Canadian Respiratory Journal : Journal of the Canadian Thoracic Society, 2008. **15**(2): p. 99-106.
- 20 872
- 21 873
- 22 874 52. Yates, D.H., *Role of exhaled nitric oxide in asthma*. Immunol Cell Biol, 2001. **79**(2): p. 178-90.
- 23 875
- 24 876 53. Lang, R.M., et al., *Recommendations for chamber quantification: a report from the American Society of Echocardiography's Guidelines and Standards Committee and the Chamber Quantification Writing Group, developed in conjunction with the European Association of Echocardiography, a branch of the European Society of Cardiology*. J Am Soc Echocardiogr, 2005. **18**(12): p. 1440-63.
- 25 877
- 26 878
- 27 879
- 28 880 54. J. Throck Watson, O.D.S., *Introduction to Mass Spectrometry: Instrumentation, Applications, and Strategies for Data Interpretation, 4th Edition*. 2007: Wiley. 862.
- 29 881
- 30 882 55. Buszewski, B., et al., *Human exhaled air analytics: biomarkers of diseases*. Biomed Chromatogr, 2007. **21**(6): p. 553-66.
- 31 883
- 32 884 56. Amann, A., et al., *Applications of breath gas analysis in medicine*. International Journal of Mass Spectrometry, 2004. **239**(2-3): p. 227-233.
- 33 885
- 34 886 57. Almstetter, M.F., P.J. Oefner, and K. Dettmer, *Comprehensive two-dimensional gas chromatography in metabolomics*. Anal Bioanal Chem, 2012. **402**(6): p. 1993-2013.
- 35 887
- 36 888 58. Das, M.K., et al., *Investigation of Gender-Specific Exhaled Breath Volatome in Humans by GCxGC-TOF-MS*. Analytical Chemistry, 2014. **86**(2): p. 1229-1237.
- 37 889
- 38 890 59. Phillips, M., et al., *Detection of an extended human volatome with comprehensive two-dimensional gas chromatography time-of-flight mass spectrometry*. PLoS One, 2013. **8**(9): p. e75274.
- 39 891
- 40 892
- 41 893 60. Gruber, B., et al., *Vacuum ultraviolet absorption spectroscopy in combination with comprehensive two-dimensional gas chromatography for the monitoring of volatile organic compounds in breath gas: A feasibility study*. J Chromatogr A, 2016. **1464**: p. 141-6.
- 42 894
- 43 895 61. Gruber, B., et al., *Breath gas monitoring during a glucose challenge by a combined PTR-QMS/GCxGC-TOFMS approach for the verification of potential volatile biomarkers*. J Breath Res, 2016. **10**(3): p. 036003.
- 44 896
- 45 897
- 46 898 62. Mellors, T.R., et al., *A new method to evaluate macaque health using exhaled breath: A case study of M. tuberculosis in a BSL-3 setting*. J Appl Physiol (1985), 2017. **122**(3): p. 695-701.
- 47 899
- 48 900 63. Phillips, M., et al., *Breath biomarkers of whole-body gamma irradiation in the Gottingen minipig*. Health Phys, 2015. **108**(5): p. 538-46.
- 49 901
- 50 902 64. Bajtarevic, A., et al., *Noninvasive detection of lung cancer by analysis of exhaled breath*. BMC Cancer, 2009. **9**: p. 348.
- 51 903
- 52 904
- 53
- 54
- 55
- 56
- 57
- 58
- 59
- 60

- 1
2
3 905 65. Lavra, L., et al., *Investigation of VOCs associated with different characteristics of breast*
4 906 *cancer cells*. Sci Rep, 2015. **5**: p. 13246.
- 5 907 66. Zhou, W., et al., *Exhaled breath online measurement for cervical cancer patients and healthy*
6 908 *subjects by proton transfer reaction mass spectrometry*. Anal Bioanal Chem, 2017. **409**(23):
7 909 p. 5603-5612.
- 8 910 67. Fernández del Río, R., et al., *Volatile Biomarkers in Breath Associated With Liver Cirrhosis —*
9 911 *Comparisons of Pre- and Post-liver Transplant Breath Samples*. EBioMedicine, 2015. **2**(9): p.
10 912 1243-1250.
- 11 913 68. O'Hara, M.E., et al., *Limonene in exhaled breath is elevated in hepatic encephalopathy*. J
12 914 *Breath Res*, 2016. **10**(4): p. 046010.
- 13 915 69. Tomasz Ligor, C.A., Anton Amann, Konrad Schwarz, Bogusaw Buszewski, Wojciech ebrowski,,
14 916 *<Comparison of Proton Transfer Reaction-Mass Spectrometry and Gas Chromatography-*
15 917 *Mass Spectrometry in Analysis of Breath Samples.pdf>*. Chem. Anal., 2009.
- 16 918 70. Darwiche, K., et al., *Bronchoscopically obtained volatile biomarkers in lung cancer*. Lung,
17 919 2011. **189**(6): p. 445-52.
- 18 920 71. Handa, H., et al., *Exhaled Breath Analysis for Lung Cancer Detection Using Ion Mobility*
19 921 *Spectrometry*. PLoS ONE, 2014. **9**(12): p. e114555.
- 20 922 72. Bessa, V., et al., *Detection of volatile organic compounds (VOCs) in exhaled breath of patients*
21 923 *with chronic obstructive pulmonary disease (COPD) by ion mobility spectrometry*.
22 924 *International Journal for Ion Mobility Spectrometry*, 2011. **14**(1): p. 7-13.
- 23 925 73. Schivo, M., et al., *Mobile instrumentation platform to distinguish airway disorders*. Journal of
24 926 *breath research*, 2013. **7**(1): p. 017113-017113.
- 25 927 74. Badjagbo, K., et al., *Direct atmospheric pressure chemical ionization-tandem mass*
26 928 *spectrometry for the continuous real-time trace analysis of benzene, toluene, ethylbenzene,*
27 929 *and xylenes in ambient air*. J Am Soc Mass Spectrom, 2009. **20**(5): p. 829-36.
- 28 930 75. Healy, R.M., et al., *A multi-year study of VOC emissions at a chemical waste disposal facility*
29 931 *using mobile APCI-MS and LPCI-MS instruments*. Environ Pollut, 2018. **232**: p. 220-228.
- 30 932 76. Linforth, R. and A.J. Taylor, *Persistence of Volatile Compounds in the Breath after Their*
31 933 *Consumption in Aqueous Solutions*. Journal of Agricultural and Food Chemistry, 2000. **48**(11):
32 934 p. 5419-5423.
- 33 935 77. Perez-Hurtado, P., et al., *Direct analysis of volatile organic compounds in foods by headspace*
34 936 *extraction atmospheric pressure chemical ionisation mass spectrometry*. Rapid Commun
35 937 *Mass Spectrom*, 2017. **31**(22): p. 1947-1956.
- 36 938 78. Berchtold, C., L. Meier, and R. Zenobi, *Evaluation of extractive electrospray ionization and*
37 939 *atmospheric pressure chemical ionization for the detection of narcotics in breath*.
38 940 *International Journal of Mass Spectrometry*, 2011. **299**(2-3): p. 145-150.
- 39 941 79. Kapishon, V., et al., *Atmospheric pressure chemical ionization mass spectrometry of pyridine*
40 942 *and isoprene: potential breath exposure and disease biomarkers*. J Breath Res, 2013. **7**(2): p.
41 943 026005.
- 42 944 80. Heaney, L.M., et al., *Real-time monitoring of exhaled volatiles using atmospheric pressure*
43 945 *chemical ionization on a compact mass spectrometer*. Bioanalysis, 2016. **8**(13): p. 1325-36.
- 44 946 81. Lê Cao, K.-A., S. Boitard, and P. Besse, *Sparse PLS discriminant analysis: biologically relevant*
45 947 *feature selection and graphical displays for multiclass problems*. BMC Bioinformatics, 2011.
46 948 **12**: p. 253-253.
- 47 949 82. Zou, H., T. Hastie, and R. Tibshirani, *Sparse Principal Component Analysis*. Journal of
48 950 *Computational and Graphical Statistics*, 2006. **15**(2): p. 265-286.
- 49 951 83. Zou H, H.T., *Regularization and Variable Selection via the Elastic Net*. Journal of the Royal
50 952 *Statistical Society*, 2005.
- 51
52
53
54
55
56
57 953
58
59
60

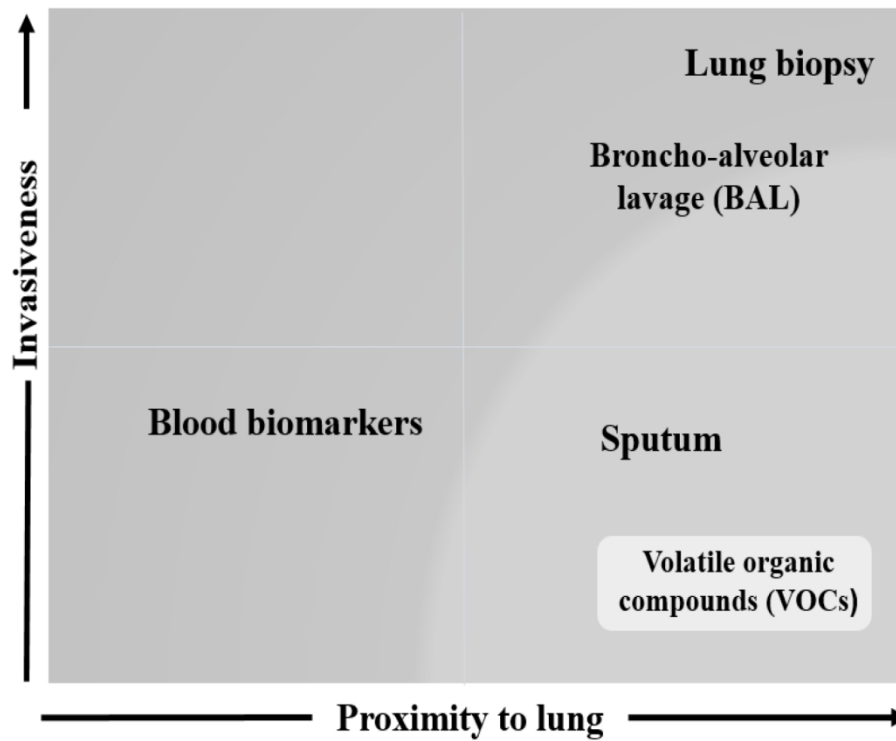


Figure (1): Relationship between lung proximity and degree of invasiveness of different lung matrices. The figure plots the level of invasiveness of various lung matrices in relation to their proximity to the lung. Given their pathological relevance, the degree of invasiveness of bronchoalveolar lavage (BAL) and lung biopsy makes them less favourable in diagnosing respiratory diseases.

144x115mm (300 x 300 DPI)

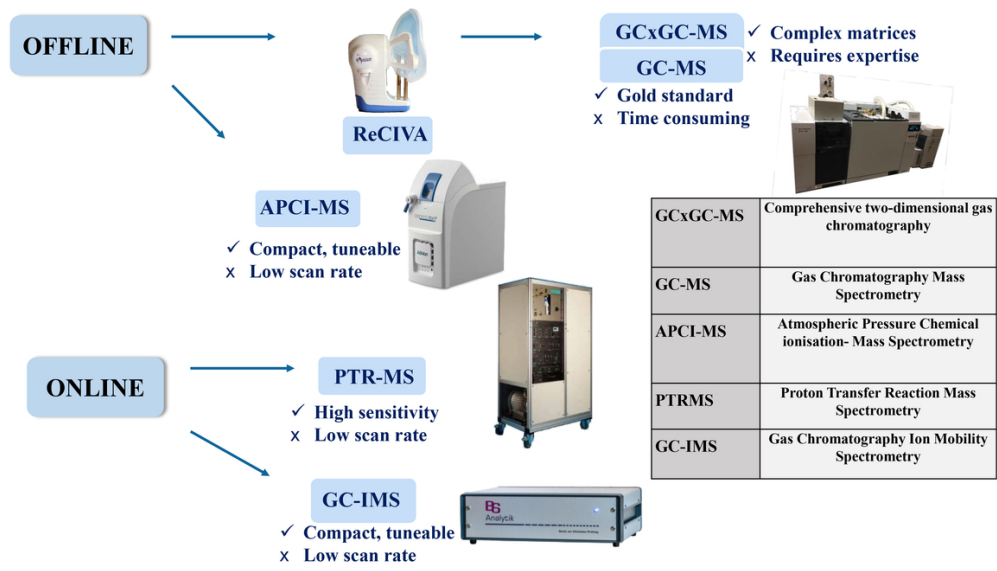
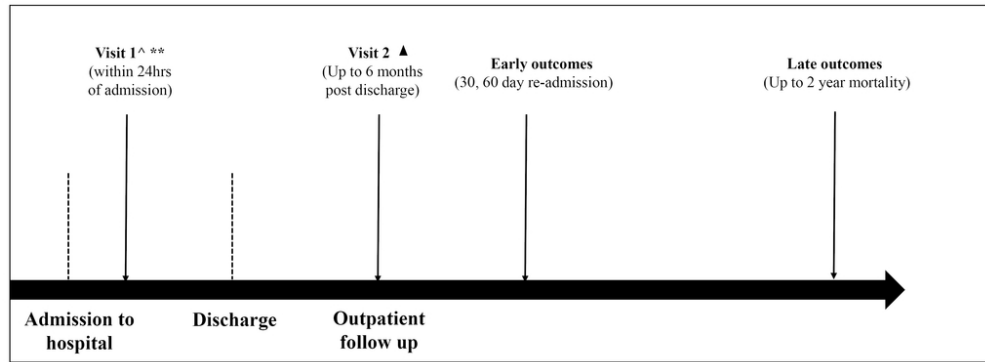


Figure (2): Multi-instrument use in breath sampling. Figure illustrates the various combinations of offline and online devices used in breath sampling and the relevant pros and cons. Offline and online technologies are used for the discovery and validation phases of the study respectively.

97x55mm (300 x 300 DPI)

BMJ Open: first published as 10.1136/bmjopen-2018-025486 on 8 March 2019. Downloaded from <http://bmjopen.bmj.com/> on April 17, 2024 by guest. Protected by copyright.



^ Following senior decision maker review
 ** Breath sampling (ReCIVA-GC-MS, ReCIVA- GCxGC-MS, GC-IMS, APCI-MS)
 ▲ Breath sampling (ReCIVA-GC-MS, ReCIVA- GCxGC-MS, PTR-MS, APCI-MS)
 ---- Following University Hospitals of Leicester streaming and clinical care pathways

Figure (3): Study flow chart. Figure outlines the patient journey from admission through to discharge and follow up. Participants with self-reported acute breathlessness presenting to University Hospitals of Leicester are recruited within 24 hours following a senior decision maker review. Breath sampling is carried out on the first visit, at recruitment, and the second visit, up to 6 months after discharge. Patients are admitted through the standard operational emergency medical streaming and care pathways at the University Hospitals of Leicester NHS Trust. Early outcomes (hospital readmission) are measured at 30 and 60 days and late outcomes (mortality) including respiratory and all-cause mortality are measured at 2 years. Assessments carried out at each time point are summarised in Table 1.

86x43mm (300 x 300 DPI)

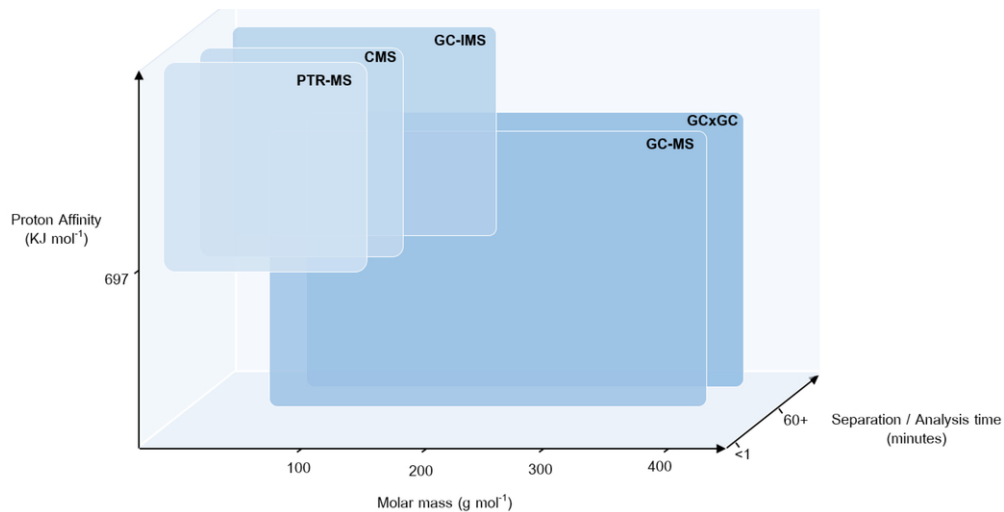


Figure (4): Multi-instrument use in breath sampling. Operational space of the analytical technologies used in EMBER for the analysis of volatile organic compounds (VOC) in exhaled breath, including proton transfer reaction-mass spectrometry (PTR-MS), atmospheric pressure chemical ionisation-mass spectrometry (CMS), gas chromatography-ion mobility spectrometry (GC-IMS), gas chromatography-mass spectrometry (GC-MS) and two-dimensional gas chromatography-mass spectrometry (GCxGC). Comparing the typical molar mass range detectable; selectivity in detection owing to the type of ionisation involved and the proton affinity of analytes; and the inclusion of a chromatographic separation affecting total time of analysis. The online technologies involving chemical ionisation (PTR-MS, CMS and GC-IMS) can be used in-clinic owing to short analysis times, but only detect lower molar mass molecules with a proton affinity higher than 697 KJ mol⁻¹. Offline chromatographic techniques (GC-MS and GCxGC) detect a wider range of compounds independent of proton affinity; however, the techniques have longer analysis times and involve sample transportation and storage.

88x44mm (300 x 300 DPI)