

ORIGINAL RESEARCH

## **Towards Optimum Reporting of Pulmonary Effectiveness**

### **Databases and Outcomes (TORPEDO): identifying a core dataset for asthma and COPD studies**

Job FM van Boven<sup>1</sup>, Sarah J Lucas<sup>2</sup>, Gary Parker<sup>3</sup>, Alan Kaplan<sup>4</sup>, Antigona Trofor<sup>5</sup>, Billie Bonevski<sup>6</sup>, Bruce J Kirenga<sup>7</sup>, Dermot Ryan<sup>8</sup>, Emilio Pizzichini<sup>9</sup>, Eric van Ganse<sup>10</sup>, Erick Wan-Chun Huang<sup>11</sup>, Evelyn Brakema<sup>12</sup>, Gillian Gould<sup>6</sup>, Janwillem Kocks<sup>13</sup>, Jennifer Alison<sup>14</sup>, Jennifer K Quint<sup>15</sup>, Joan B Soriano<sup>16</sup>, John Hurst<sup>17</sup>, Kamran Siddiqi<sup>18</sup>, Katherine Boydell<sup>19</sup>, Marc Miravittles<sup>20</sup>, Mario Alberto Flores-Valdez<sup>21</sup>, Marise Kasteleyn<sup>12</sup>, Mark FitzGerald<sup>22</sup>, Melanie Boeckmann<sup>23</sup>, Michael Chaiton<sup>24</sup>, Miguel Roman Rodriguez<sup>25</sup>, Muralidhar Kulkarni<sup>26</sup>, Nicolas Roche<sup>27</sup>, Niels Chavannes<sup>12</sup>, Nikolaos G Papadopoulos<sup>28,29</sup>, Panagiotis Behrakis<sup>30</sup>, Sarah Dennis<sup>14</sup>, Shalini Bassi<sup>31</sup>, Siân Williams<sup>32</sup>, Toby M Maher<sup>33</sup>, Trishul Siddharthan<sup>34</sup>, Veena Kamath<sup>26</sup>, Katia MC Verhamme<sup>35</sup> on behalf of REG/GACD TORPEDO collaborators.

1. University of Groningen, University Medical Center Groningen, Netherlands
2. Respiratory Effectiveness Group, Ely, UK
3. Global Alliance for Chronic Diseases (GACD), London, UK
4. Department of Family and Community Medicine, University of Toronto, Toronto, ON, Canada
5. Universitatea de Medicina si Farmacie Grigore T. Popa Iasi, Romania
6. The University of Newcastle, Australia
7. Makerere University, Uganda
8. The University of Edinburgh, United Kingdom

9. Federal University of Santa Catarina, Brazil
10. University of Lyon, France
11. Woolcock Institute of Medical Research, Australia
12. Leiden University Medical Center, Netherlands
13. General Practitioners Research Institute, Netherlands
14. The University of Sydney, Australia
15. Imperial College London, United Kingdom
16. Instituto de Investigación Hospital Universitario de la Princesa, Spain
17. University College London, United Kingdom
18. University of York, United Kingdom
19. University of New South Wales, Australia
20. El Vall d'Hebron Barcelona Hospital, Spain
21. Center for Research and Applied Technology of Jalisco, Guadalajara, Mexico
22. Vancouver Coastal Health, Canada
23. Bielefeld University, School of Public Health, Department of Environment and Health, Germany
24. Cancer Care Ontario, Canada
25. Servei de Salut de les Illes Balears, Spain
26. Manipal University, India
27. Hôpitaux Universitaires Paris Centre, Hôpital Cochin, AP-HP and Université de Paris, Paris, France
28. Division of Infection, Immunity & Respiratory Medicine, University of Manchester, Manchester, UK
29. Allergy Department, 2nd Pediatric Clinic, National Kapodistrian University of Athens, Athens, Greece

30. Hellenic Cancer Society, Greece
31. Public Health Foundation of India, India
32. International Primary Care Respiratory Group (IPCRG), United Kingdom
33. National Heart and Lung Institute, Imperial College, London, UK
34. Johns Hopkins School of Public Health, United States
35. Department of Medical Informatics, Erasmus MC, Rotterdam, Netherlands

**Key Words:** Asthma, COPD, Epidemiology.

**Corresponding author:** Job van Boven. University of Groningen, University Medical Center Groningen, Netherlands. Email: [jobvanboven@gmail.com](mailto:jobvanboven@gmail.com)

**Tables/Figures:** 2 tables / 3 figures (plus 2 supplementary tables)

**REG/GACD TORPEDO collaborators who participated in the modified Delphi exercise:**

Ana Quintana, Universidad Peruana Cayetano Heredia, Peru; Brusselle Guy, Ghent University Hospital, Belgium; Chin Kook Rhee, The Catholic University of Korea, Korea; Emmanuel Guindon, McMaster University, Canada; Faisal Yunus, Universitas Indonesia, Indonesia; Francisco Gonzalez-Salazar, Universidad de Monterrey, Mexico; Guillermo Paraje, Universidad Adolfo Ibáñez, Chile; Guy Marks, University of New South Wales, Australia; Hana Ross, University of Cape Town, South Africa; Helen Reddel, The University of Sydney, Australia; Ian Glaspole, Alfred Health, Australia; Ivan Kimuli, Makerere University Lung Institute, Uganda; Jiapeng Lu, National Center for Cardiovascular Diseases, China; John Britton, University of Nottingham, United Kingdom; Maria Kathia Cardenas, Universidad Peruana Cayetano Heredia, Peru; Michael Schatz, Kaiser Permanente, USA; Naveen Puttaswamy, Sri Ramachandra University, India; Nicole

Vellios, University of Cape Town, South Africa; Olena Tigova, Catalan Institute of Oncology, Spain; Pawan Tiwari, All India Institute of Medical Sciences, India; Raana Zahid, The Initiative, Pakistan; Robert Schwartz, University of Toronto, Canada; Steve Turner, University of Aberdeen, United Kingdom; Sudeepa Khanal, HERD International (Health Research and Social Development Forum), Nepal; Thys van der Molen, University Medical Center Groningen, Netherlands; Tomi Kovacevic, Institut za Plucne Bolesti Vojvodine, Serbia.

**Funding:** This study was supported by the Respiratory Effectiveness Group, an international, investigator-led, not-for-profit, real-life respiratory research and advocacy initiative (<http://www.regresearchnetwork.org>) and the Global Alliance for Chronic Diseases ([www.gacd.org](http://www.gacd.org)).

## Abstract

**Purpose:** There remains a need for a standardized dataset for respiratory studies to accelerate data collection, improve research efficiency and aid the sharing, merging and comparison of datasets. This TORPEDO (Towards Optimum Reporting of Pulmonary Effectiveness Databases and Outcomes) project aimed to develop a checklist of optimum and minimum variables for asthma and chronic obstructive pulmonary disease (COPD) research.

**Methods:** A 3-phase modified Delphi survey was conducted: in phase 1, an expert panel generated a list of variables, in phase 2 a Delphi panel selected the minimum variables (>66% agreement) for any design and in phase 3 they were asked to select a minimum set for specific study designs.

**Results:** In phase 1 the expert panel (n=22) proposed 224 variables. In phase 2, voting by 64 participants resulted in consensus (>66% agreement) for 18 variables and partial agreement (50-66%) for 44 variables, following this, 5 technical variables (e.g. date of test) were removed. In phase 3, 34 members of the Delphi panel completed voting; consensus was reached for 13 variables for retrospective asthma studies and 34 for prospective asthma studies. For COPD, there were 16 variables for retrospective studies and 37 for prospective studies. Gender, asthma/COPD exacerbations and patient-reported outcomes were the only variables with 100% agreement for both asthma and COPD studies.

**Conclusion:** The proposed list of minimally required variables will allow the assessment of current data sources for their utility in asthma and COPD studies, facilitate the merging of datasets, aid standardization of data collection and improve research efficiency.

## Introduction

Randomized controlled trials (RCTs) are considered the gold standard to assess efficacy, and, to a limited extent, the safety of a drug or non-pharmacological asthma or chronic obstructive pulmonary disease (COPD) treatment. However, because of the stringent methodology adopted with strict inclusion and exclusion criteria seen in RCTs, the relatively small sample size and short duration of follow-up, observational studies and pragmatic trials are required to provide additional information on the effectiveness and safety of a drug when used in real-life circumstances.<sup>1-3</sup> Electronic health care databases, claims databases and drug and/or patient registries are important data sources for real-life studies, which has recently been underscored by regulatory authorities such as the European Medicines Agency (EMA) and the UK's National Institute for Health and Care Excellence (NICE).<sup>4</sup> With the increasing digitalization of healthcare data collection, the number of sources available for observational studies is growing exponentially, allowing research centers to conduct multinational, multi-database studies.

Heterogeneity of database structures, and differences in disease and drug coding complicates the conduct of these studies.<sup>5</sup> These concerns about heterogeneity not only hold for databases, but also for the choice and definition of outcomes which makes a comparison between studies - not to mention pooling of data - difficult. The importance of choosing realistic and clinically meaningful outcomes has been described for various clinical domains including the field of respiratory research. Indeed, in 2008, the *European Respiratory Journal* published the recommendations of an (American Thoracic Society/ European Respiratory Society (ATS/ERS) task force on appropriate outcomes for COPD pharmacological trials, but no priority-analysis was performed.<sup>6</sup>

The Respiratory Effectiveness Group (REG, <https://regresearchnetwork.org/>) is an international network of respiratory experts that aims to set standards and best practices to improve real-world respiratory research. REG has made efforts to set standards for real-life respiratory research in general, but not yet at the level of specific variables.<sup>7</sup> Thus, there is still an urgent need to standardize outcomes and develop a core dataset for asthma and COPD studies to accelerate data collection, improve research efficiency, replicability and transparency and create the possibility to share, compare and merge datasets around the globe to permit further analysis.<sup>8</sup>

The Global Alliance on Chronic Diseases (GACD) is a network of the world's biggest public research funding agencies and funds joint programs on chronic diseases. Together, the members of the alliance represent over 90% of public health research funding worldwide. GACD aims to facilitate research collaboration on chronic disease globally with a focus on collaborations between low- and middle- income countries and vulnerable populations within high-income countries.<sup>9</sup> The network coordinates and supports research activities that address the prevention and treatment of chronic non-communicable diseases, on a global scale. The work of GACD on the standardization of outcomes and development of dictionaries for chronic diseases is crucial to reach these objectives. As part of a joint effort between the REG and GACD, the TORPEDO (Towards Optimum Reporting of Pulmonary Effectiveness Databases and Outcomes) project aimed to develop a checklist with optimum and minimum required variables for research in asthma and COPD.

## Methods

### *Study design*

This study followed the process of a three-phased modified Delphi process<sup>10</sup> as outlined in Figure 1. For this Delphi survey, various partners involved in respiratory research (clinicians, regulators, respiratory societies, patient organizations, researchers) were asked to complete a checklist asking what they considered to be optimum and minimum required variables for asthma and COPD research followed by a prioritization process of multiple rounds to reach consensus on the list of variables. The method was based on previous studies that aimed to reach consensus on a core set of items required for proper reporting of datasets.<sup>11,12</sup> A point of difference to the classic Delphi approach was that, in phase 1, a smaller selected expert panel was consulted instead of the full Delphi panel to first expand the initial list with additional variables for an optimum asthma/COPD dataset. From phase 2 onwards, a full Delphi panel was established, and a prioritization process followed. The process was conducted between March 2018 and April 2019. All phases are further detailed in the next sections.

This study involved no patients; participants were all professionals involved in respiratory research who are members of the REG or GACD networks and consented to participate in the Delphi exercise. The study received ethical approval from the Anonymised Data Ethics and Protocol Transparency (ADEPT) Committee (approval number ADEPT0921).

### *Outcomes*

We defined a 'minimum' research dataset to be the minimum amount of clinical data variables to be collected to establish the effects of an intervention related to the prevention



and management of asthma and COPD. An ‘optimum’ set would be the larger set of clinical data variables that could be collected where time and other related resources related to their collection and analysis were less constrained. This will provide researchers with a benchmark of which variables to select given their research goals, anticipated design and resources available.

## **Phase 1**

### *Identifying an optimum asthma/COPD dataset*

The authors started with an initial set of variables that had been used in previous asthma/COPD studies, which were selected at the authors’ discretion. This list was circulated among a panel of respiratory experts, who could complement it with relevant variables according to their perspective. A requirement for this expert panel was that it had representation of: (i) researchers and clinicians with expertise in asthma, COPD, allergy, primary care, epidemiology and/or health economics, (ii) each continent and (iii) representatives of the major international asthma and COPD guidelines committees (Global Initiative for Asthma (GINA), Global Initiative for Chronic Obstructive Lung Disease (GOLD)). These members were all recruited through the REG (<https://www.regresearchnetwork.org/>). Once the list was expanded, the full list of all variables was discussed at the REG Summit (London, UK, 2017) and, if required, some variables were merged or re-categorized.

## **Phase 2**

In phase 2, the expert panel from phase 1 was expanded to the formal Delphi panel with more global representatives, focusing on representatives from low- and middle-income countries. These additional members were recruited through the GACD respiratory

disease program. By combining the networks of REG and GACD we aimed for a representative Delphi panel where we considered the following requirements:

- High-income and low-income country representation
- Good geographical representation (all World Health Organisation regions represented, i.e. Africa, Americas, South East Asia, Europe, Eastern Mediterranean and Western Pacific)
- Wide range of researchers and clinicians with expertise in asthma, COPD, allergy, primary care, epidemiology and/or health economics
- Representation from asthma and COPD guideline committees (GINA, GOLD)

*Voting and endorsing of variables to reduce the list to a 'minimum' dataset*

In phase 2 of this project, all variables identified in phase 1 were presented to the Delphi panel through an online voting platform (Survey Monkey; [www.surveymonkey.com](http://www.surveymonkey.com)). For each variable, the Delphi panel members were requested to indicate if they considered the variable an absolutely required (minimum) variable for any asthma/COPD research (no matter the design of the study, as specific design was addressed in phase 3). Variables that reached at least >66% agreement between respondents were included in the list and were moved to the next Delphi round (phase 3). Variables with partial agreement (50-66%) were further discussed at REG (Paris, France, 2018) and GACD meetings (Sao Paulo, Brazil, 2018) and if deemed relevant by the majority, included for the next Delphi round (phase 3). The 66% and 50-66% cut-off criteria were based on a previous modified Delphi exercise in severe asthma<sup>12</sup>.

### **Phase 3**

### *Prioritization of the minimum dataset per disease and study design*

In this round, a prioritization of the variables identified in phase 2 was made. The Delphi panel was provided with a survey sent out through the same online voting platform. The Delphi members were presented with the list of variables identified during phase 2. Respondents were asked to indicate which variables were a minimum requirement for the study designs below:

- Prospective clinical asthma (field) study with original data collection
- Prospective clinical COPD (field) study with original data collection
- Retrospective asthma database study
- Retrospective COPD database study

As for phase 2, variables were included in the list if at least >66% agreement between respondents was reached. This exercise resulted in four final lists of minimally required variables for four types of respiratory studies.

## Results

An overview of variable selection is provided in Figure 2 and detailed in the following subsections by phase.

### Phase 1

#### *Expert panel and optimum variable list*

After the suggestions of the expert panel (N=22) and REG meeting discussions, an initial list of 224 variables was generated. This list consisted of all variables deemed relevant for asthma/COPD research and is presented in Supplementary Table 1 by domain. This list was presented to the Delphi panel consulted in phase 2.

### Phase 2

#### *Delphi panel*

The characteristics of the Delphi panel are presented in Table 1. There was more representation from high-income than lower/middle-income countries. The majority of Delphi members had (clinical) academic positions and most indicated they had primarily in the asthma or COPD expertise, but there was also strong representation from epidemiology experts.

#### *List of minimum respiratory variables (uncategorized)*

A total of 64 Delphi members participated in phase 2. By means of online voting on the initial set of 224 variables, immediate consensus (ie >66% agreement) was reached for 18 variables (8%). These variables are indicated by the dark grey fill in Supplementary Table 1. Partial agreement (50-66%) was reached for 44 variables (20%). These variables are

indicated by the light grey fill in Supplementary Table 1. The latter were discussed at the REG (Paris) and GACD (Sao Paulo) 2018 meetings and most were deemed relevant but some only for specific studies (eg retrospective database only or prospective only) or only in specific type of studies. After phase 2, a list of 62 variables remained. After discussion, generic “technical variables” (date of test, units) were removed resulting in a set of 57 variables to be considered in phase 3.

### **Phase 3**

In total, 64 invitations to participate in phase 3 were sent out to the Delphi panel of which 39 members replied (61%). Of these, 5 people did not select any of variables and were hence excluded from the analysis, thus the final participants were 34 (53%).

#### *Final list of minimum variables by disease and study design*

After voting in this phase, from the total of 57 variables remaining from phase 2, 13 core variables were considered for retrospective asthma studies and 34 for prospective asthma studies. For COPD, this was 16 variables for retrospective studies versus 37 for prospective studies. In prospective studies, the additional variables required were mainly biomarkers, lung function measurements, and variables providing further information concerning healthcare utilization and reason for visit. Of note, gender, asthma/COPD exacerbations and a relevant patient-reported outcome (Asthma Control Questionnaire (ACQ), COPD Assessment Test (CAT)) were the only variables with 100% agreement for both asthma and COPD studies. The exact percentages of agreement per variable is provided in Supplementary Table 2. The final list of variables per disease (asthma, COPD) and study design (retrospective, prospective) is provided in Table 2 and Figure 3.

## Discussion

We used a modified Delphi exercise to determine a recommended core dataset of minimum variables for asthma and COPD studies. Following the identification of the full list of potential asthma/COPD variables by a panel of 22 REG experts, 64 participants completed phase 2 to determine the minimally recommended variables, 34 of these completed phase 3 to determine the minimal variables per study type. There was generally good agreement concerning the variables needed for retrospective studies, regardless of whether the study was investigating asthma or COPD; similarly, for prospective studies, the variables required were very similar for both COPD and asthma studies. For prospective studies, where new data collection would be possible, the participants felt that a greater number of variables should be required than for retrospective studies. In prospective studies, the additional variables required were mainly biomarkers, lung function measurements, and variables giving further information concerning healthcare utilization and reason for visit. The fewer required variables in retrospective studies likely reflects the experience of the participants who were aware of the limitations of retrospective data sources. Indeed, the accuracy and regularity of recording of variables might influence a participant's decision on whether that variable should be included, and even its utility.

To facilitate database studies, but also for existing longitudinal cohorts, there is a need to determine whether databases contain the minimum required variables to adequately conduct the study. Often in observational database studies, particularly for rare events and conditions, it is beneficial to include data from multiple data sources to increase the power. Also, the ability to pool data from multiple datasets allows a more heterogeneous population of patients to be included from different cultural backgrounds. The

Uncovering and Noting Long-term Outcomes in COPD and asthma to enhance Knowledge (UNLOCK) initiative, created in 2010, aimed to determine the minimum required dataset for observational studies in asthma and COPD management in primary care, which is an important step to enable pooling of data from multiple datasets.<sup>13</sup> Indeed, the UNLOCK group subsequently found that a key challenge in using real-world data from across different countries and regions was the lack of comparability between datasets, with not all of them containing all the required data variables or having the same definitions of the same variable.<sup>14</sup> A difference with our study is that in our study a structured Delphi approach was used (compared to a focus group discussion in the UNLOCK study) and that we expanded our research beyond primary care and existing datasets.

Data reuse can be severely hampered when datasets are not standardized. Therefore, it is important to ensure the minimum required variables are included in new (thus prospective) data collection.<sup>15</sup> The Core Outcome Measures in Effectiveness Trials (COMET) Initiative also aims to develop standardized sets of outcomes, and these sets should be considered the minimum variables that should be measured and reported in clinical trials.<sup>15</sup> Importantly, such lists did not yet exist for asthma and/or COPD and needed to be developed for use beyond clinical trials only. The development of core datasets is also important to understand the implementation of interventions. Where we have standardized datasets, we are more likely be able to understand differences between centers as a result of true differences rather than differences in variables.

Several groups have previously considered defining core variables or have developed a list of minimum required variables using a Delphi approach. A similar style of Delphi exercise was used to determine the minimum variables required when creating the International Severe Asthma Registry (ISAR), which reached a consensus on 95 variables

including patient demographics, medical history, patient-reported outcomes, diagnostic information, clinical characteristics and physician-reported outcomes.<sup>12</sup> While the ATS/ERS Task Force on “Outcomes for COPD pharmacological trials: from lung function to biomarkers” did not determine a list of minimum variables they did highlight certain variables that COPD trials should include, such as lung function variables other than forced expiratory volume in 1 second (FEV1) (e.g. forced vital capacity (FVC), inspiratory capacity to total lung capacity ratio and measures of dyspnoea) and the frequency of exacerbations.<sup>6</sup> Information on COPD exacerbations was included in both our retrospective and prospective COPD lists, and some lung function measures other than FEV1 were included on the prospective COPD study list. An UNLOCK study, combining data from primary care across Europe to determine the prevalence of comorbidities in COPD patients, and their impact on health status and COPD, required datasets to contain the following variables: age, gender, FEV1 and ideally FEV1/FVC, CAT or Clinical COPD Questionnaire score, Medical Research Council (MRC) dyspnea score, body mass index, smoking status, education level and comorbidities (Heart disease, hypertension, diabetes, depression and asthma).<sup>16</sup> These required variables are comparable to those included in our list, but there are some differences (in comorbidities and requirement for MRC score), which are mainly due to the specific requirements of the UNLOCK study. Indeed, the list produced here provides purely the core recommended variables; in practice the minimum variables required for a particular study will depend on the specific question to be addressed.

Pooling of data, not only from clinical trials but also with regard to real-world observational studies becomes increasingly more important, and harmonization of data by means of a common data model not only optimizes data extraction but also data pooling.<sup>17</sup> Ideally, the



dictionary used by the common data model should also include the minimum variables as suggested in our retrospective asthma and COPD lists.

One of the strengths of this study is that the Delphi panel (n=64) comprised of a wide range of participants from: high- and low and-middle income countries, different geographic regions, different areas of expertise and different professional backgrounds. However, only 34 participants (53%) participated in both phase 2 and 3, and the reason to abstain for phase 3 is not known. This is, however, a common challenge in online Delphi studies and given we have no reason to assume that biased exclusion occurred (“missing at random”), we feel that it did not significantly impact the validity of our results.

While it has been determined that some variables such as the full date of birth were required, often local regulatory/legal rules such as General Data Protection Regulation (GDPR) might not allow inclusion of potential identifying patient information.<sup>18</sup> Therefore, other formats or related variables may need to be used, such as only including year of birth or age at the date of enrolment.

This list is valid at the present time but may require updating in the future as the data required changes over time with increased knowledge giving a better understanding of risk factors, improved diagnostic criteria and clinical management. For example, FEV1 measurements are currently key in diagnosing COPD, however, it is hoped in the future there may be validated biomarkers for COPD, which may need to be included in a future list of minimum required dataset.<sup>6</sup>

Lastly, there remains debate around the precise definition of some disease terms, for example exacerbations, and for standardization across studies it is important that the same definition has been used.

Determined by a global Delphi panel of 64 participants, these proposed minimum required variables will facilitate the assessment of current data sources for their potential utility for use in asthma and COPD studies. It provides a basis to aid the standardization of data collection and improve research efficiency, replicability and transferability. Determining these minimal variables is an important step in facilitating the sharing, comparing and merging of datasets.

## Potential conflicts of interest

A. Kaplan is a member of advisory board or speakers bureau for Astra Zeneca, Behring, Boehringer Ingelheim, Covis, GSK, Pfizer, Purdue, Merck Frosst, Novartis, NovoNordisc, Sanofi, Teva, Trudel. J. Kocks reports grants, personal fees and non-financial support from AstraZeneca, grants, personal fees and non-financial support from Boehringer Ingelheim, grants and personal fees from Chiesi Pharmaceuticals, grants, personal fees and non-financial support from GSK, grants and personal fees from Novartis, grants from MundiPharma, grants from TEVA, outside the submitted work; and Janwillem Kocks holds 72.5% of shares in the General Practitioners Research Institute. Dr. Kocks reports grants and non-financial support from AstraZeneca, grants and non-financial support from Boehringer Ingelheim, grants from Chiesi Pharmaceuticals, grants and non-financial support from GSK, grants from Novartis, grants from MundiPharma, grants from TEVA, from null, outside the submitted work; and Janwillem Kocks holds 72.5% of shares in the General Practitioners Research Institute. N. Roche reports grants and personal fees from Boehringer Ingelheim, Novartis, Pfizer and personal fees from Teva, GSK, AstraZeneca, Chiesi, Sanofi, Trudell, Zambon. E. Van Ganse has received consultancy fees from Pelyon Ltd, for unrelated research projects. J. Hurst has received support to attend meeting, and payment for education and advisory work (personally and to his employer, UCL) from pharmaceutical companies that make medicines to treat COPD. M. Romàn-Rodríguez reports personal fees and grants from AstraZeneca, Boehringer Ingelheim, GSK, Mundipharma, Novartis, Teva, and Trudell Medical International in the last 3 years. M. Miravittles has received speaker fees from AstraZeneca, Boehringer Ingelheim, Chiesi, Cipla, Menarini, Rovi, Bial, Sandoz, Zambon, CSL Behring, Grifols and Novartis, consulting fees from AstraZeneca, Boehringer

Ingelheim, Chiesi, GlaxoSmithKline, Bial, Gebro Pharma, Kamada, CSL Behring, Laboratorios Esteve, Ferrer, Mereo Biopharma, Verona Pharma, TEVA, Spin Therapeutics, pH Pharma, Novartis, Sanofi and Grifols and research grants from GlaxoSmithKline and Grifols. N. Papadopoulos reports personal fees from Novartis, personal fees from Nutricia, personal fees from HAL, personal fees from MENARINI/FAES FARMA, personal fees from SANOFI, personal fees from MYLAN/MEDA, personal fees from BIOMAY, personal fees from AstraZeneca, personal fees from GSK, personal fees from MSD, personal fees from ASIT BIOTECH, personal fees from Boehringer Ingelheim, grants from Gerolymatos International SA, grants from Capricare, outside the submitted work. K. Verhamme works for a research group that received/receives unconditional grants from Yamanouchi, Pfizer/Boehringer Ingelheim, GSK, Novartis none of which are related to the content of this manuscript. All other authors report no other conflicts of interest in relation to this work.

## References

1. Roche N, Reddel HK, Agusti A, et al. Integrating real-life studies in the global therapeutic research framework. *Lancet Respir Med*. 2013;1(10):e29-30.
2. Roche N, Reddel H, Martin R, et al. Quality standards for real-world research. Focus on observational database studies of comparative effectiveness. *Ann Am Thorac Soc*. 2014;11 (Suppl 2):S99-S104.
3. Roche N, Anzueto A, Anticevich SB, et al. The importance of Real-Life research in Respiratory Medicine: Manifesto of the Respiratory Effectiveness Group Endorsed by the International Primary Care Respiratory Group and the World Allergy Organization. *Eur Respir J*. 2019;54:1901511.
4. Pacurariu A, Plueschke K, McGettigan P, et al. Electronic healthcare databases in Europe: descriptive analysis of characteristics and potential for use in medicines regulation. *BMJ Open*. 2018; 8:e023090.
5. Trifiro G, Coloma PM, Rijnbeek PR, et al. Combining multiple healthcare databases for postmarketing drug and vaccine safety surveillance: why and how? *J Intern Med*. 2014; 275(6):551-561.
6. Cazzola M, MacNee W, Martinez FJ, et al. Outcomes for COPD pharmacological trials: from lung function to biomarkers. *Eur Respir J*. 2008;31(2):416-469.
7. Roche N, Campbell JD, Krishnan JA, et al. Quality standards in respiratory real-life effectiveness research: the REal Life EVidence AssessmeNt Tool (RELEVANT): report

from the Respiratory Effectiveness Group-European Academy of Allergy and Clinical Immunology Task Force. *Clin Transl Allergy*. 2019;9:20.

8. Weeks J, Pardee R. Learning to Share Health Care Data: A Brief Timeline of Influential Common Data Models and Distributed Health Data Networks in U.S. Health Care Research. *EGEMS (Wash DC)*. 2019;7(1):4.

9. Riddell MA, Edwards N, Thompson SR, et al. Developing consensus measures for global programs: lessons from the Global Alliance for Chronic Diseases Hypertension research program. *Global Health*. 2017;13(1):17.

10. Pill J. The Delphi method: substance, context, a critique and an annotated bibliography. *Socio-Economic Planning and Science*. 1971;5, 57–71.

11. Husereau D, Drummond M, Petrou S, et al. Consolidated Health Economic Evaluation Reporting Standards (CHEERS)--explanation and elaboration: a report of the ISPOR Health Economic Evaluation Publication Guidelines Good Reporting Practices Task Force. *Value Health*. 2013;16(2):231-250.

12. Bulathsinhala L, Eleangovan N, Heaney LG, et al. Development of the International Severe Asthma Registry (ISAR): A Modified Delphi Study. *J Allergy Clin Immunol Pract*. 2019;7(2):578-588.

13. Chavannes N, Stallberg B, Lisspers K, et al. UNLOCK: Uncovering and Noting Long-term Outcomes in COPD to enhance knowledge. *Prim Care Respir J*. 2010;19(4):408.

14. Cragg L, Williams S, van der Molen T, Thomas M, Correia de Sousa J, Chavannes NH. Fostering the exchange of real world data across different countries to answer primary care research questions: an UNLOCK study from the IPCRG. *NPJ Prim Care Respir Med*. 2018;28(1):8.
15. The Lancet Respiratory Medicine. Maximising the data potential in COPD research. *Lancet Respir Med*. 2017;5(7):535.
16. Stallberg B, Teixeira P, Blom C, et al. The prevalence of comorbidities in COPD patients, and their impact on health status and COPD symptoms in primary care patients: a protocol for an UNLOCK study from the IPCRG. *NPJ Prim Care Respir Med*. 2016;26:16069.
17. Voss EA, Makadia R, Matcho A, et al. Feasibility and utility of applications of the common data model to multiple, disparate observational health databases. *J Am Med Inform Assoc*. 2015;22(3):553-564.
18. van Veen EB. Observational health research in Europe: understanding the General Data Protection Regulation and underlying debate. *Eur J Cancer*. 2018;104:70-80.

**Table 1:** Characteristics of the Delphi panel participants in phase 2 (N=64)

<b>Characteristic</b>	<b>N (%)</b>
Male	40 (63%)
Female	24 (38%)
<b>Income</b>	
High-income country	42 (66%)
Low and middle income country	22 (34%)
<b>World Health Organization region</b>	
African region	6 (9%)
Americas	13 (20%)
South East Asia (inc. India, Indonesia, Nepal)	8 (13%)
Europe	26 (41%)
Eastern Mediterranean	0 (0%)
Western Pacific (inc. Australia, China, Korea)	11 (17%)
<b>Main workplace</b>	
Academia	42 (66%)
Hospital	11 (17%)
General practice	3 (5%)
Industry	2 (3%)
Regulator	1 (2%)
Other	5 (8%)
<b>Core area of expertise (self-indicated)</b>	
Asthma	9 (14%)
COPD	12 (19%)
Primary care	8 (13%)
Pediatrics	2 (3%)
Allergy	1 (2%)
ILD/IPF	2 (3%)
Epidemiology	12 (19%)
Health economics	6 (9%)
Other	12 (19%)

ILD: interstitial lung disease; IPF: idiopathic pulmonary fibrosis

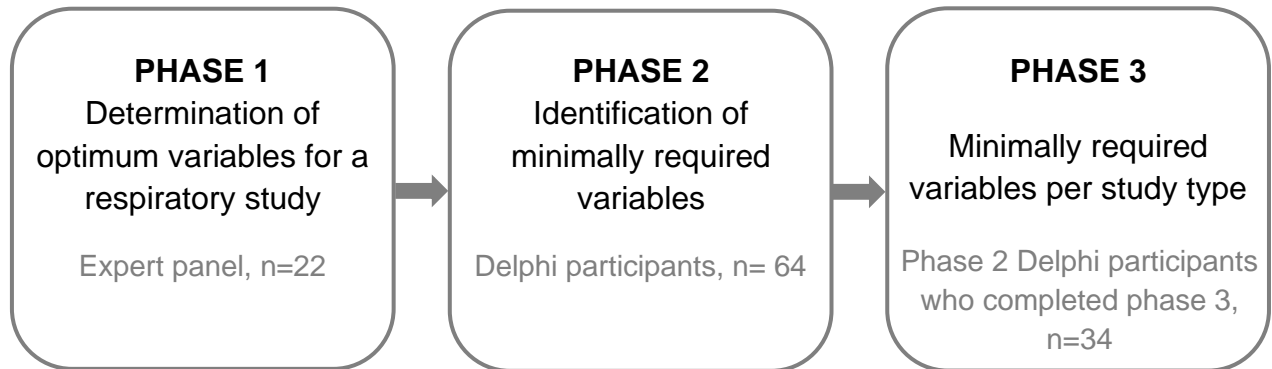


**Table 2:** Final list of minimum variables, indicated by **V**, by disease and study design

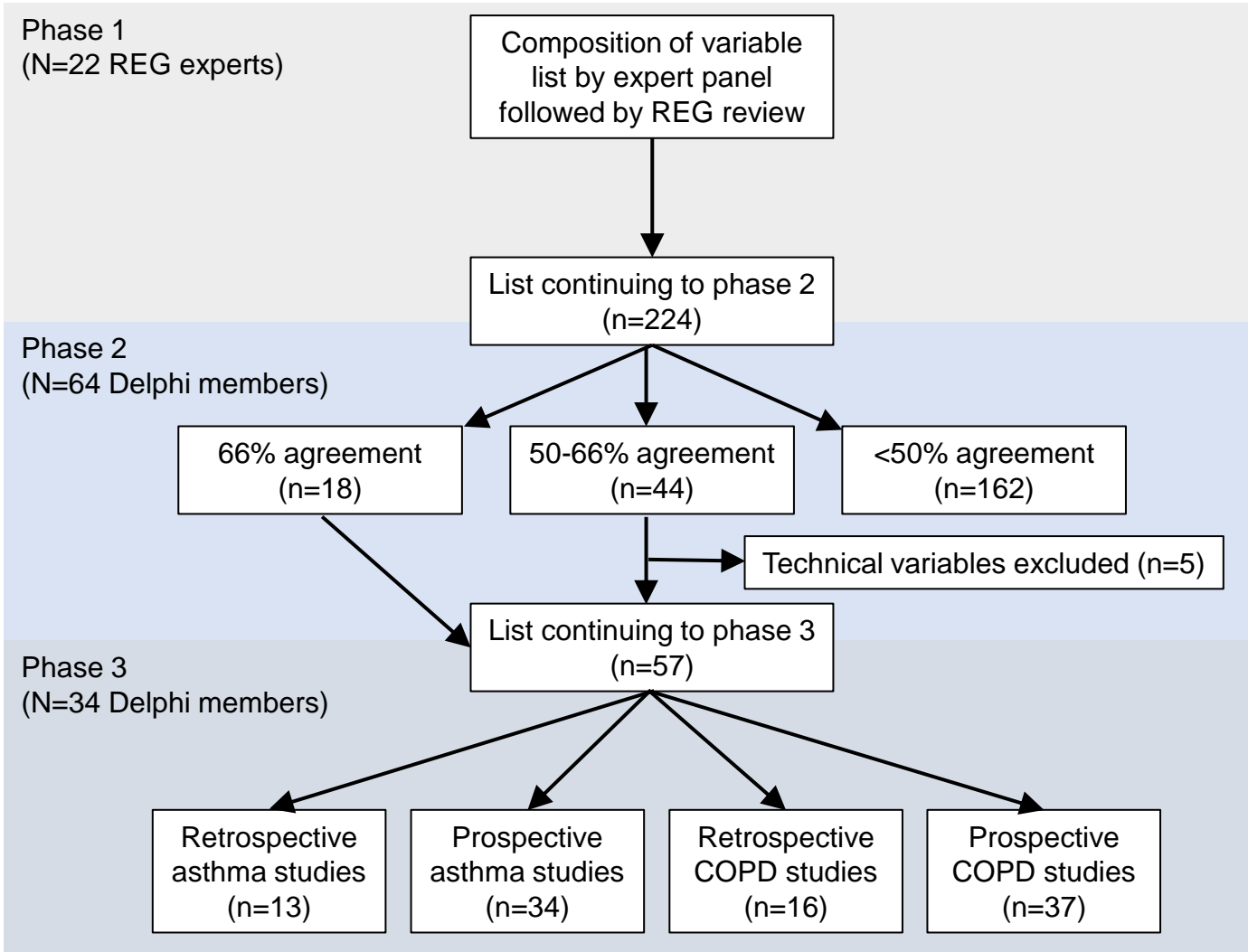
	<b>Suggested format</b>	<b>Asthma Retrospective</b>	<b>Asthma Prospective</b>	<b>COPD Retrospective</b>	<b>COPD Prospective</b>
<b>Demographics</b>					
Date of birth	DD-MM-YYYY	V	V	V	V
Gender	M/F	V	V	V	V
Geographical location	Country/region		V		V
<b>Clinical</b>					
Height	Meters		V		V
Weight	Kg				V
Body Mass Index	Kg/m <sup>2</sup>		V		V
Environmental exposures	Job type		V		V
Smoking status	Current/former/ never	V	V	V	V
Pack years	Packyears	V	V	V	V
<b>Biomarkers</b>					
FBC with differentiation	Eosinophils, neutrophils etc.		V		V
Immunoglobulin E	U		V		
<b>Lung function</b>					
FEV1_prebronchodilation	Liters		V		V
FEV1_postbronchodilation	Liters		V		V
FEV1% predicted	%		V	V	V
FVC	Liters		V		V
<b>Pollution</b>					
Allergen exposure	Yes or no		V		
Occupational exposure	"				V
<b>Healthcare utilization</b>					
Primary care consultation	Number/date of visits	V	V	V	V
Reason Primary Care Visit	Diagnosis		V		V
Secondary care visit	Date and number		V		V
Hospitalisation	Date and number	V	V	V	V
Reason hospitalisation	Diagnosis		V		V
ER admission	Date and number	V	V	V	V
Reason ER admission	Diagnosis		V		V
ICU stay	Date and number		V		V
<b>Medication</b>					
Generic drug name	Name	V	V	V	V
Frequency of use	Date and number		V		V
<b>Comorbidities</b>					
Diabetes Mellitus	Yes or No			V	V
Rhinitis	"	V	V		
Lung cancer	"			V	V
Bronchiectasis	"				V
Asthma	"			V	V
COPD	"	V	V		
Cardiovascular disease	"			V	V
<b>Mortality</b>					
Mortality	Life status (dead/alive)	V	V	V	V
Cause of death	Diagnosis		V		V
Primary Cause of death	Diagnosis		V		V
Date of death	Date	V	V	V	V
<b>Exacerbations</b>					
Asthma exacerbations	Date and number	V	V		
COPD exacerbations	"			V	V
Other severe respiratory episodes	"		V		V
<b>Patient reported outcomes</b>					
ACQ	Score		V		
CAT	"				V

ACQ: asthma control questionnaire; CAT: COPD Assessment Test; ER: emergency room; FBC: full blood count; FEV1: forced expiratory volume in 1 second; FVC: forced vital capacity; ICU: intensive care unit.

**Figure 1: Overview of the Delphi phases**

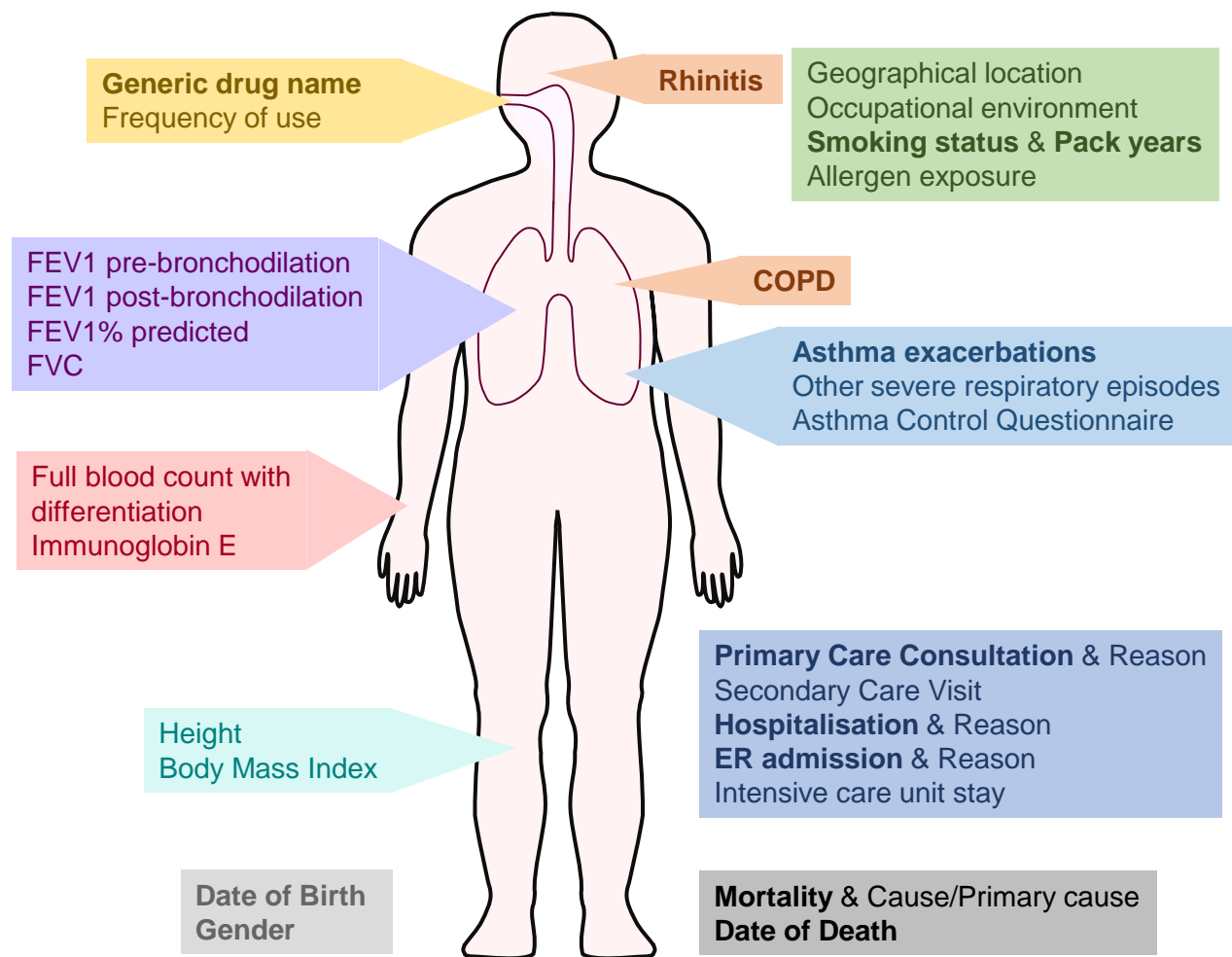


**Figure 2:** Variable flow diagram. *REG: Respiratory Effectiveness Group, COPD: Chronic Obstructive Pulmonary Disease*



**Figure 3:** Overview of REG/GACD core dataset for prospective asthma (A) and COPD (B) studies, with variables for retrospective studies indicated in bold. COPD: chronic obstructive pulmonary disease; ER: emergency room; FEV1: forced expiratory volume in 1 second; FVC: forced vital capacity.

A) Asthma variables



B) COPD variables

