The international primary ciliary dyskinesia cohort (iPCD Cohort): methods and first results

Myrofora Goutaki1,2, Elisabeth Maurer1, Florian S. Halbeisen1, Israel Amirav3, Angelo Barbato4 on behalf of the PCD Italian Consortium, Laura Behan5, Mieke Boon6, Carmen Casaulta2 on behalf of the Swiss PCD Group, Annick Clement7 on behalf of the French Reference Centre for Rare Lung Diseases, Suzanne Crowley8, Eric Haarman9, Claire Hogg10, Bulent Karadag11, Cordula Koerner-Rettberg12, Margaret W. Leigh13 on behalf of the Genetic Disorders of Mucociliary Clearance Consortium, Michael R. Loebinger14, Henryk Mazurek15, Lucy Morgan16, Kim G. Nielsen17, Heymut Omran18, Nicolaus Schwerk19, Sergio Scigliano20, Claudius Werner18, Panayiotis Yiallouros21, Zorica Zivkovic22,23, Jane S. Lucas5 and Claudia E. Kuehni1

@ERSpublications

The iPCD Cohort offers a unique opportunity to study PCD in an international retrospective cohort of >3000 patients http://ow.ly/rn0m304Jgsu

Cite this article as: Goutaki M, Maurer E, Halbeisen FSThe international primary ciliary dyskinesia cohort (iPCD Cohort): methods and first results. Eur Respir J 2017; 49: 1601181 [https://doi.org/10.1183/13993003.01181-2016].

ABSTRACT    Data on primary ciliary dyskinesia (PCD) epidemiology is scarce and published studies are characterised by low numbers. In the framework of the European Union project BESTCILIA we aimed to combine all available datasets in a retrospective international PCD cohort (iPCD Cohort).

We identified eligible datasets by performing a systematic review of published studies containing clinical information on PCD, and by contacting members of past and current European Respiratory Society Task Forces on PCD. We compared the contents of the datasets, clarified definitions and pooled them in a standardised format.

As of April 2016 the iPCD Cohort includes data on 3013 patients from 18 countries. It includes data on diagnostic evaluations, symptoms, lung function, growth and treatments. Longitudinal data are currently available for 542 patients. The extent of clinical details per patient varies between centres. More than 50% of patients have a definite PCD diagnosis based on recent guidelines. Children aged 10–19 years are the largest age group, followed by younger children (<9 years) and young adults (20–29 years).

This is the largest observational PCD dataset available to date. It will allow us to answer pertinent questions on clinical phenotype, disease severity, prognosis and effect of treatments, and to investigate genotype-phenotype correlations.

This article has supplementary material available from erj.ersjournals.com

Received: June 15 2016 | Accepted after revision: Sept 27 2016

Copyright ©ERS 2017. This ERJ Open article is open access and distributed under the terms of the Creative Commons Attribution Non-Commercial Licence 4.0.
Affiliations: 1Institute of Social and Preventive Medicine, University of Bern, Bern, Switzerland. 2Paediatric Respiratory Medicine, Children’s University Hospital of Bern, University of Bern, Bern, Switzerland. 3Dept of Pediatrics, Bar Ilan University, Ramat Gan, Israel. 4Primary Ciliary Dyskinesia Centre, Dept of Women’s and Children’s Health (SDB), University of Padova, Padova, Italy. 5Primary Ciliary Dyskinesia Centre, NIHR Southampton Respiratory Biomedical Research Unit, University of Southampton and University Hospital Southampton NHS Foundation Trust, Southampton, UK. 6Dept of Paediatrics, University Hospital Gasthuisberg, Leuven, Belgium. 7Paediatric Pulmonary Dept. Trouseau Hospital APHP, Sorbonne Universities and Pierre et Marie Curie University, Paris, France. 8Unit for Paediatric Heart, Lung and Allergic Diseases, Rikshospitalet, Oslo, Norway. 9Dept of Pulmonary Medicine, VU University Medical Center, Amsterdam, The Netherlands. 10Dept of Paediatrics, Primary Ciliary Dyskinesia Centre, Royal Brompton and Harefield Foundation Trust, London, UK. 11Dept of Paediatric Pulmonology, Marmara University, School of Medicine, Istanbul, Turkey. 12Dept of Paediatric Pneumology, University Children’s Hospital of Ruhr University Bochum, Bochum, Germany. 13Dept of Pediatrics, UNC School of Medicine, Chapel Hill, NC, USA. 14Host Defence Unit, Royal Brompton and Harefield NHS Foundation Trust, London, UK. 15Dept of Pneumology and Cystic Fibrosis, Institute of Tuberculosis and Lung Disorders, Rabka-Zdroj, Poland. 16Dept of Respiratory Medicine, Concord Hospital Clinical School, University of Sydney, Sydney, Australia. 17Danish PCD Centre Copenhagen, Paediatric Pulmonary Service, Copenhagen University Hospital, Copenhagen, Denmark. 18Dept of General Paediatrics and Adolescent Medicine, University Hospital Muenster, Muenster, Germany. 19Clinic for Paediatric Pulmonology, Allergology and Neonatology, Hannover Medical School, Hannover, Germany. 20Centro Respiratorio, Hospital de Niños Ricardo Gutierrez, Buenos Aires, Argentina. 21Medical School, University of Cyprus, Nicosia, Cyprus. 22Children’s Hospital for Lung Diseases and TB, Medical Centre “Dr Dragisa Misovic”, Belgrade, Serbia. 23Faculty of Pharmacy, University Business Academy, Novi Sad, Serbia.

Correspondence: Claudia E. Kuehni, Institute of Social and Preventive Medicine, University of Bern, Finkenhubelweg 11, Bern 3012, Switzerland. E-mail: claudia.kuehni@ispm.unibe.ch

Introduction
Primary ciliary dyskinesia (PCD) is a rare heterogeneous disease; genetic mutations cause functional and/ or structural defects of cilia [1, 2]. This results in chronic upper and lower respiratory disease, such as progressive chronic suppurative lung disease, chronic serous otitis media (glue ear) and chronic rhinosinusitis [3]. Nearly half of PCD patients have situs inversus [4] and some have other heterotaxic syndromes or congenital heart defects [5, 6]. PCD affects about 1 in 10 000 people [2, 7]. As in most orphan diseases, research has focused on identifying the responsible genes, describing pathophysiological mechanisms and improving diagnostic tests [8–11]. Clinical characteristics have been described mainly in small case series and literature reviews. A recent meta-analysis found a wide range in reported prevalence of clinical manifestations of PCD [12]. Earlier reports on PCD in children suggested a relatively benign long-term course. This has been questioned by recent publications, which demonstrated that many adult patients develop severe lung disease with chronic Pseudomonas infection, become oxygen dependent and eventually require lung transplantation [3, 13–16]. Although growth and lung function are important predictors of severity and prognosis in many lung diseases [17], data on PCD is scant, showing contradictory results [13, 15, 18–23]. There are no age-standardised data on mortality and little is known about factors influencing long-term prognosis. For instance, it is not clear how mortality and lung function are influenced by different ultrastructural and genetic defects.

Although 7% of the population suffers from one of about 7000 rare diseases [24], research is scarce for these disorders. Research on PCD and many other rare diseases cannot utilise routine data such as mortality and hospital episode statistics because most rare diseases do not have a dedicated International Classification of Diseases revision 10 code and can therefore not be identified in routine statistics. The low numbers of patients in each individual centre call for collaborative research and international studies are essential.

Aware of these issues, clinicians and scientists with a strong interest in PCD formed an international focus group in 2006 and took a number of initiatives to advance PCD research. The first Task Force on PCD (2006–2009) of the European Respiratory Society (ERS) included 26 countries [1, 7, 25]. This European network, under the framework of the European Union’s Seventh Framework Programme (EU FP7) project BESTCILIA (Better Experimental Screening and Treatment for Primary Ciliary Dyskinesia), joined forces with the North American Genetic Disorders of Mucociliary clearance Consortium. Two work packages of support statement: The development of the iPCD Cohort has been funded from the European Union’s Seventh Framework Programme under EG-GA 35404 BESTCILIA: Better Experimental Screening and Treatment for Primary Ciliary Dyskinesia. Primary ciliary dyskinesia research at ISPBM Bern is also funded by national funding from the Lung Leagues of Bern, St Gallen, Vaud, Ticino and Valais and the Milena Carvajal Pro-Kartagener Foundation. The researchers participate in the network of COST Action BEAT-PCD: Better Evidence to Advance Therapeutic options for PCD (BM 1407). Funding information for this article has been deposited with the Open Funder Registry.

Conflict of interest: Disclosures can be found alongside this article at erj.ersjournals.com
BESTCILIA aimed to improve the availability of international datasets for PCD patients: 1) by setting up a prospective international PCD registry [26], to allow standardised data collection in the future, and 2) by identifying and combining available data on PCD in a retrospective international cohort study. This article describes the aims and methods of the international PCD cohort (iPCD Cohort) and outlines how the data can be accessed for future research.

Aims of the iPCD Cohort
The iPCD Cohort assembles available datasets with clinical and diagnostic data from PCD patients worldwide to answer pertinent questions on clinical phenotype, disease severity, prognosis and effect of treatments in patients with this rare multiorgan disease.

This combined international dataset allows investigation of PCD epidemiology in a large international study population in order to: 1) describe the spectrum of clinical phenotypes and disease severity in PCD patients by age, sex and time period of diagnosis; 2) describe short-term and long-term prognosis of PCD, looking at important outcomes such as growth, lung function and respiratory failure, bacterial colonisation, hearing loss, fertility, and mortality; and 3) identify predictors of long-term outcomes such as age at diagnosis, clinical phenotype, ultrastructural defects, genotype and clinical care.

Methods
Study design
The iPCD Cohort is a retrospective international cohort, combining available data on PCD from national or local registries and clinical or diagnostic databases. All participating centres delivered retrospectively collected data; new centres joining the iPCD Cohort in the future can also participate with retrospectively collected data.

Identification of eligible datasets
Systematic literature search
In order to identify published data and eligible datasets we performed a systematic literature search for studies describing clinical manifestations in patients with PCD [12]. We searched the online databases PubMed, Embase and Scopus for studies published since 1980, without restrictions on language or study design, including studies containing information on clinical manifestations of 10 or more PCD patients. We also checked the reference lists of articles to find additional studies. We identified 52 different studies and invited all corresponding authors to participate in the iPCD Cohort.

Other sources
Using the database set up in a previous European PCD survey [1, 5, 7] and personal contacts from the ERS PCD Task Force, we contacted all clinicians who had reported treating PCD patients and asked them to collaborate. This included physicians from Europe, North and South America, the Middle East, and Australia.

Data pooling and standardisation
Overall, diagnostic definitions have changed over the years. Cut-offs for diagnostic test results depend on laboratory conditions, techniques and equipment used (e.g. the reference range for ciliary beat frequency is dependent on the temperature at which analyses are made). Definitions of many clinical manifestations also vary between countries and centres. We observed considerable heterogeneity in content and format between the delivered datasets. For this reason, and to obtain comparable data for pooled analyses, we developed a standardised dataset. We discussed the content and format of this dataset with all collaborators, who include experts on PCD diagnostics and clinical care, and with PCD patient representatives, participating at BESTCILIA meetings. For diagnostic tests results, in particular, we chose the use of categorical variables in order to achieve standardisation and avoid misclassification at the time of analysis. Content and format of the variables were tailored to the availability of retrospective data, as identified by examination of the received datasets and the prospective international PCD registry [26]. This facilitates the close collaboration between the two datasets for future research.

Diagnostic subgroups
PCD diagnostics have evolved rapidly in recent years [27]. Initially, diagnosis was based on the Kartagener symptoms triad [28] and on transmission electron microscopy (TEM) findings. Then light microscopy and later high-frequency video microscopy (VM) were introduced in the diagnostic algorithm. Currently, recommendations include a combination of TEM, VM, nasal nitric oxide (nNO) and genetic testing [11], but availability of tests differs between countries [27]. Therefore, not all PCD patients have been diagnosed according to current recommendations. The iPCD Cohort includes patients diagnosed since 1964, when diagnostic criteria were different. As a result of this, we defined three different diagnostic subgroups based on the results of the available tests. We used the recent guidelines of the ERS PCD Diagnostics Task Force [11] to
define “definite PCD” as hallmark TEM findings and/or bi-allelic PCD genetic mutation. “Probable PCD” was defined when patients had abnormal VM findings and/or low nNO, using a cut-off of 77 nL·min⁻¹, as previously published [29]. All patients who had negative or ambiguous test results, or had not been tested so far, were defined as having “clinical PCD diagnosis”. These patients are followed up and treated as PCD at the collaborating centres based on a combination of several of the following features: situs anomalies, persistent cough, persistent rhinitis, chronic or recurrent upper or lower respiratory infections and history of neonatal respiratory symptoms in term infants, based on consensus statements and guidelines available to age [1, 11].

What information is collected
The iPCD Cohort includes retrospectively collected patient data on the following 11 thematic categories (table 1): 1) general information, 2) results of diagnostic tests, 3) baseline characteristics, 4) growth and lung function, 5) clinical manifestations, 6) therapy, 7) microbiology, 8) imaging, 9) surgical interventions, 10) neonatal period, and 11) family history. Details on all variables included in the standardised dataset and information on the coding of variables are included in online supplementary table S1.

Infrastructure for data delivery and data rights
The iPCD Cohort is hosted at the Institute of Social and Preventive Medicine at the University of Bern, Switzerland. Research is performed in close collaboration with all data contributors. We have set up a safe information technology platform (Sharepoint) for uploading already available anonymised datasets. We organised telephone conferences to offer extra support during the initial steps of the data delivery, and to discuss questions and comments from the contributing partners. To further facilitate data delivery in the standardised dataset for new partners we created a web-based platform, using the software Research Electronic Data Capture (REDCap) developed at Vanderbilt University (Nashville, TN, USA), which is

<table>
<thead>
<tr>
<th>Type of data</th>
<th>Collected variables</th>
<th>Contributing centres</th>
</tr>
</thead>
<tbody>
<tr>
<td>General information</td>
<td>Patient and centre identifiers, and demographic data (e.g. date of birth, sex, ethnicity, date of last follow-up, date last known alive)</td>
<td>All centres</td>
</tr>
<tr>
<td>Diagnostics</td>
<td>Diagnostic status of PCD, date of diagnosis, dates, and results of diagnostics test performed (e.g. nNO testing, VM, TEM, genetic testing)</td>
<td>All centres</td>
</tr>
<tr>
<td>Baseline characteristics</td>
<td>Information on PCD defects (e.g. situs anomalies, congenital heart defects, brain cilia dysfunction, retinitis pigmentosa, renal problems, fertility problems) and relevant comorbidities</td>
<td>All centres</td>
</tr>
<tr>
<td>Lung function</td>
<td>Lung function date, somatometric measurement, spirometric measurements and other lung function tests (e.g. multiple breath washout)</td>
<td>All centres except those in Northern America</td>
</tr>
<tr>
<td>Clinical manifestations</td>
<td>Date of clinical visit, information on symptoms and signs (e.g. cough, upper and lower respiratory infections, hearing impairment), exercise limitation and active or passive smoking within the last 3 months</td>
<td>AR, AU, CH, CY, DE2, DE3, DK, FR, IL, IT, NL, NO, PL, RS, TR, UK2, UK3</td>
</tr>
<tr>
<td>Therapy</td>
<td>Date of prescription, information on inhaled and nasal medication (e.g. bronchodilators, inhaled corticosteroids, nasal sprays), antibiotics prescription, oxygen therapy, mechanical ventilation and physiotherapy</td>
<td>AU, CH, CY, DE3, IL, IT, NO, PL, RS, TR, UK3</td>
</tr>
<tr>
<td>Microbiology</td>
<td>Date of microbiology examination, type of sample and isolated bacteria</td>
<td>AR, BE, CH, CY, DE2, DE3, DK, NO, PL, RS, TR, UK2, UK3</td>
</tr>
<tr>
<td>Imaging</td>
<td>Date of chest imaging examination, type of chest imaging, information on findings (e.g. bronchiectases, atelectases, infiltrations) and information on other types of imaging</td>
<td>BE, CH, CY, DE2, DE3, FR, IL, IT, NO, RS, TR, UK2, UK3</td>
</tr>
<tr>
<td>Surgical interventions</td>
<td>Date and type of surgery</td>
<td>AR, AU, BE, CH, CY, DE3, IT, NL, NO, RS, TR</td>
</tr>
<tr>
<td>Neonatal period</td>
<td>Symptoms and signs of neonatal period (e.g. neonatal respiratory distress syndrome, neonatal cough, neonatal rhinitis, need for intensive care and ventilation)</td>
<td>AU, BE, CH, CY, DE2, DE3, FR, IL, IT, NL, NO, RS, TR, UK3</td>
</tr>
<tr>
<td>Family history</td>
<td>Number of siblings, PCD affected siblings and other family members, patient ID of PCD affected family members, and relationship of affected family members</td>
<td>AR, AU, BE, CH, CY, DE2, IL, IT, NO, RS, TR, UK3</td>
</tr>
</tbody>
</table>

nNO: nasal nitric oxide; VM: light or high-frequency video microscopy; TEM: transmission electron microscopy. #: AR: Argentina; AU: Australia; BE: Belgium; CH: Switzerland; CY: Cyprus; DE1: Bochum, Germany; DE2: Muenster, Germany; DE3: Hannover, Germany; DK: Denmark; FR: France; IL: Israel; IT: Italy; NL: the Netherlands; NO: Norway; PL: Poland; RS: Serbia; TR: Turkey; UK1: Paediatric Pulmonology Dept, Brompton, UK; UK2: Adult Pulmonology Dept, Brompton, UK; UK3: Southampton, UK. #: mandatory data for all centres.
widely used in the academic research community [30], and allows data entry and extraction in various formats. Detailed instructions explain how data can be entered. The REDCap environment is completely secure and each contributor only has access to the data of their own centre. We collaborated closely with centres that already had national registries and large retrospective datasets, to help them recode the data into the standardised format, and we performed quality controls of the delivered data (e.g. checks for coding errors and plausibility of entered values and dates). Each partner was responsible for ensuring that the delivered data were anonymous and in accordance with the national/local data protection laws. More information on data protection can be found in the online supplementary material under Data protection/Ethics. We drafted agreements for data delivery and publication (see online supplementary material), which leave all rights with data contributors.

Data analysis
For each planned analysis, all eligible data are validated, cleaned and standardised. We identify outliers and implausible values, and if necessary we contact contributors to resolve any issues encountered. Based on the research question, we use a one- or two-stage random effect individual patient data meta-analysis approach. As PCD diagnostics has evolved quickly and not all patients have diagnostic information that is up to current standards, all analyses will be stratified by level of diagnostic certainty, using the diagnostic subgroups described previously. Whenever relevant, we will perform additional sensitivity analysis including only patients with definite PCD, based on the recent guidelines of the ERS PCD Diagnostics Task Force [11], and will compare these results with the overall analysis results.

Funding
The setting up of the iPCD Cohort (salaries, consumables and equipment) was funded by the EU FP7 project BESTCILIA (http://bestcilia.eu) and several Swiss funding bodies, including the Lung Leagues of Bern, St Gallen, Vaud, Ticino and Valais and the Milena Carvajal Pro-Kartagener Foundation. Data collection and management at each site was funded according to local arrangements. Most participating researchers and data contributors participate in the COST Action BEAT-PCD: Better Evidence to Advance Therapeutic options for PCD (BM 1407; www.beatpcd.org). Infrastructure is provided for free by the University of Bern, where the data are pooled and stored.

Results
Current dataset
The iPCD Cohort is an ongoing effort where new centres can still join and participating centres can add data for new patients. As of April 2016, data from 21 single centres or consortia from 18 countries had been
contributed. The pooled dataset contained data on 3013 patients (figure 1). Some countries which have a national reference centre for all paediatric and adult PCD patients (Cyprus, Denmark and France) or a national PCD registry (Italy and Switzerland) contributed their national dataset. Other countries contributed data from their consortia of several centres (Israel, USA and Canada) or datasets from single-centre studies (all remaining centres). The number of patients per centre ranged from 10 to 436 (table 2). Two paediatric centres contributed datasets only with patients <18 years and one with adult patients (≥20 years old). Five paediatric centres additionally included a few young adults. The remaining 13 centres submitted datasets that included both paediatric and adult patients with an age range from 0 to 77 years. The percentage of males ranged from 38% to 68% among datasets. General information, results of diagnostic tests and baseline characteristics (Modules 1–3) are in the basic mandatory dataset, and thus available from all centres and for all 3013 patients or the majority of them (tables 1 and 3). In addition to these baseline data, several centres have also contributed data to the other modules (table 3). Data on growth are currently available for more than 1500 patients and on lung function for more than 1000 patients. Additional data, such as microbiology or perinatal history, have been delivered by several groups (table 1 and figure 1) and for several hundreds of patients (table 3). Longitudinal data have been contributed for 542 patients from 10 countries (table 2), with a follow-up period ranging from 2 to 20 years.

Patient characteristics
Characteristics of the 3013 patients included in the iPCD Cohort are described in table 4. 49% (1490) were male. 71% of patients were followed up in European centres (39% in Western Europe), 9% in Western Asia, 17% in America and 4% in Australia. More than half of the patients (table 4) have a definite PCD diagnosis based the recent guidelines of the ERS PCD Diagnostics Task Force [11]. 14% of patients were defined as probable PCD based on their diagnostic test results. In 30% of patients the test results were ambiguous or the diagnostic algorithm had not been concluded and the diagnosis was based on clinical grounds based on existing consensus or guidelines. Median (range) current age of patients was 18 (1–92) years (figure 2). Children aged between 10 and 19 years were the largest age group (38%), followed by younger children (0–9 years) and young adults (20–29 years) at 17% and 18%, respectively. The percentage of patients in the remaining age groups had a declining trend from 10% for patients aged 30–39 years to 4% for patients aged >60 years.

<table>
<thead>
<tr>
<th>Country</th>
<th>Principal investigator</th>
<th>Patients n</th>
<th>Age range years#</th>
<th>Males %</th>
<th>Type of data delivered</th>
<th>Data richness¶</th>
</tr>
</thead>
<tbody>
<tr>
<td>Argentina</td>
<td>S. Scigliano</td>
<td>101</td>
<td>6–57</td>
<td>42</td>
<td>Cross-sectional</td>
<td>Intermediate</td>
</tr>
<tr>
<td>Australia</td>
<td>L. Morgan</td>
<td>109</td>
<td>0–76</td>
<td>60</td>
<td>Longitudinal</td>
<td>High</td>
</tr>
<tr>
<td>Belgium</td>
<td>M. Boon</td>
<td>82</td>
<td>0–69</td>
<td>45</td>
<td>Cross-sectional</td>
<td>High</td>
</tr>
<tr>
<td>Cyprus</td>
<td>P. Yiallouros</td>
<td>31</td>
<td>0–66</td>
<td>48</td>
<td>Longitudinal</td>
<td>High</td>
</tr>
<tr>
<td>Denmark</td>
<td>K.G. Nielsen</td>
<td>120</td>
<td>0–70</td>
<td>48</td>
<td>Longitudinal</td>
<td>Intermediate</td>
</tr>
<tr>
<td>France</td>
<td>A. Clement</td>
<td>337</td>
<td>0–69</td>
<td>52</td>
<td>Longitudinal</td>
<td>High</td>
</tr>
<tr>
<td>Germany (Bochum)</td>
<td>C. Koerner-Reittberg</td>
<td>64</td>
<td>0–27</td>
<td>40</td>
<td>Cross-sectional</td>
<td>High</td>
</tr>
<tr>
<td>Germany (Muenster)</td>
<td>H. Omran</td>
<td>436</td>
<td>3–75</td>
<td>52</td>
<td>Cross-sectional</td>
<td>Basic</td>
</tr>
<tr>
<td>Germany (Hannover)</td>
<td>N. Schwerk</td>
<td>37</td>
<td>0–39</td>
<td>68</td>
<td>Longitudinal</td>
<td>High</td>
</tr>
<tr>
<td>Israel</td>
<td>I. Amirav</td>
<td>210</td>
<td>0–60</td>
<td>56</td>
<td>Cross-sectional</td>
<td>High</td>
</tr>
<tr>
<td>Italy</td>
<td>Italian PCD group</td>
<td>331</td>
<td>0–73</td>
<td>50</td>
<td>Cross-sectional</td>
<td>Intermediate</td>
</tr>
<tr>
<td>The Netherlands</td>
<td>E. Haarman</td>
<td>82</td>
<td>3–69</td>
<td>50</td>
<td>Longitudinal</td>
<td>Intermediate</td>
</tr>
<tr>
<td>Norway</td>
<td>S. Crowley</td>
<td>23</td>
<td>0–18</td>
<td>65</td>
<td>Longitudinal</td>
<td>High</td>
</tr>
<tr>
<td>Poland</td>
<td>H. Mazurek</td>
<td>105</td>
<td>1–22</td>
<td>56</td>
<td>Cross-sectional</td>
<td>High</td>
</tr>
<tr>
<td>Serbia</td>
<td>Z. Zivkovic</td>
<td>10</td>
<td>6–19</td>
<td>45</td>
<td>Longitudinal</td>
<td>High</td>
</tr>
<tr>
<td>Switzerland</td>
<td>Swiss PCD group</td>
<td>108</td>
<td>3–70</td>
<td>51</td>
<td>Longitudinal</td>
<td>High</td>
</tr>
<tr>
<td>Turkey</td>
<td>B. Karadag</td>
<td>37</td>
<td>3–21</td>
<td>43</td>
<td>Cross-sectional</td>
<td>Intermediate</td>
</tr>
<tr>
<td>UK (Paediatric Pulmonology Dept, Brompton)</td>
<td>C. Hogg</td>
<td>116</td>
<td>1–18</td>
<td>47</td>
<td>Cross-sectional</td>
<td>Basic</td>
</tr>
<tr>
<td>UK (Adult Pulmonology Dept, Brompton)</td>
<td>M. Loebinger</td>
<td>152</td>
<td>20–76</td>
<td>38</td>
<td>Cross-sectional</td>
<td>Basic</td>
</tr>
<tr>
<td>UK (Southampton)</td>
<td>J. Lucas</td>
<td>104</td>
<td>0–68</td>
<td>49</td>
<td>Longitudinal</td>
<td>Intermediate</td>
</tr>
<tr>
<td>USA/Canada</td>
<td>Genetic Diseases of Mucociliary Clearance Consortium</td>
<td>418</td>
<td>0–77</td>
<td>44</td>
<td>Cross-sectional</td>
<td>Basic</td>
</tr>
</tbody>
</table>

iPCD Cohort 3013 0–77 50

# at time of data delivery; ¶: semiquantitative measure based on the number of delivered variables.
Discussion

The iPCD Cohort is the largest PCD dataset available to date. It includes data from 3013 PCD patients from 18 countries; more than half of them have a definite PCD diagnosis using current standards. The main strength of the iPCD Cohort is the large number of patients it includes. This makes it a valuable

<table>
<thead>
<tr>
<th>Type of data</th>
<th>Patients</th>
</tr>
</thead>
<tbody>
<tr>
<td>General information</td>
<td>3013 [100]</td>
</tr>
<tr>
<td>Diagnostics</td>
<td>3013 [100]</td>
</tr>
<tr>
<td>nNO</td>
<td>1021 [33]</td>
</tr>
<tr>
<td>TEM</td>
<td>1913 [62]</td>
</tr>
<tr>
<td>VM</td>
<td>1088 [35]</td>
</tr>
<tr>
<td>Genetics</td>
<td>276 [9]</td>
</tr>
<tr>
<td>Baseline characteristics*</td>
<td>2286 [74]</td>
</tr>
<tr>
<td>Growth</td>
<td>1609 [53]</td>
</tr>
<tr>
<td>Lung function</td>
<td>1042 [34]</td>
</tr>
<tr>
<td>Clinical manifestations</td>
<td>1352 [44]</td>
</tr>
<tr>
<td>Therapy</td>
<td>843 [27]</td>
</tr>
<tr>
<td>Microbiology</td>
<td>732 [24]</td>
</tr>
<tr>
<td>Imaging</td>
<td>526 [17]</td>
</tr>
<tr>
<td>Neonatal period</td>
<td>1179 [38]</td>
</tr>
</tbody>
</table>

Table 3: Available data from different modules of the standardised dataset of 3013 primary ciliary dyskinesia (PCD) patients in the international PCD cohort (iPCD Cohort) (April 2016)

Data are presented as n (%). nNO: nasal nitric oxide; TEM: transmission electron microscopy; VM: light or high-frequency video microscopy. *: mainly data on situs anomalies and congenital heart disease.

Table 4: Basic characteristics of the 3013 primary ciliary dyskinesia (PCD) patients included in the international PCD cohort (iPCD Cohort) (April 2016)

Patients 3013 [100]
Sex
Male 1490 [49]
Female 1523 [51]
Location of country of residence#
Australia 109 [4]
Northern Europe 515 [17]
Western Europe 1146 [39]
Eastern Europe 105 [4]
Southern Europe 341 [11]
Western Asia 278 [9]
Northern America 418 [14]
Southern America 101 [3]
Diagnostic information
Definite PCD diagnosis¶ 1718 [56]
Probable PCD diagnosis* 420 [14]
Clinical diagnosis only 875 [30]
Age years
0–9 517 [17]
10–19 1156 [38]
20–29 553 [18]
30–39 316 [10]
40–49 205 [7]
50–59 137 [5]
≥60 129 [4]

Data are presented as n (%). #: based on geographical region definitions of the United Nations Statistic Division (August 2016) [Northern Europe: Denmark, Norway, UK; Western Europe: Belgium, France, Germany, Switzerland, the Netherlands; Eastern Europe: Poland; Southern Europe: Italy, Serbia; Western Asia: Cyprus, Israel, Turkey; Northern America: Canada, USA; Southern America: Argentina]; ¶: defined as hallmark PCD electron microscopy findings and/or bi-allelic gene mutation identified based on the European Respiratory Society PCD Diagnostics Task Force guidelines [11]; *: abnormal light or high-frequency video microscopy finding and/or low (≤77 nL·min⁻¹) nasal nitric oxide value.
source for research, especially in the field of rare diseases where even large national reference centres may follow a small number of patients. In addition to the basic characteristics and diagnostic test results, which are available from all centres and for more than 3000 patients, the iPCD Cohort includes data from over 1000 patients on neonatal symptoms, growth, lung function and clinical characteristics, and from several hundred patients on other topics such as microbiology, imaging and therapy (e.g. growth, lung function, clinical manifestations) from a number of centres; these are higher numbers of patients than were ever previously included in publications on these topics. What is important is that most centres have included almost complete datasets in the different thematic modules, in cases where they have contributed data to them. Another strength is the multinational nature of the iPCD Cohort. It offers the possibility, unique for rare diseases, to assess differences in PCD characteristics between countries and ethnic groups, and variations in healthcare (diagnostics and treatment). The iPCD Cohort includes patients of all age groups and although the paediatric population is larger, this offers the opportunity to study the course of the disease through life. Preliminary analyses on growth, nutritional status and lung function from the iPCD Cohort have been presented at international conferences, illustrating the quality of the dataset and the feasibility of future studies using the iPCD Cohort. The user-friendly and safe environment of REDCap for the standardised entry of the data is an additional strength. Data contributors can easily type in their data following simple instructions and access it any time with password-protected accounts (box 1). In addition, they can easily export their dataset for their own analyses.

Some limitations are inherent to the retrospective nature of the iPCD Cohort. For instance, there remains some heterogeneity in definitions among centres. This is not an issue for standardised measurements, such as height, weight, spirometry and microbiology/radiology test results, but is more relevant for clinical signs. Reference ranges of diagnostic test results vary, and depend greatly on the expertise, conditions, equipment and protocol used at each centre. For standardisation with comparable values, we used categorical coding for numerical values (e.g. frequency of ciliary beating has been coded as normal/decreased/increased). All the items collected in the cohort and their coding were extensively discussed and agreed among collaborators during the development period, taking into account existing national registries and the prospective international PCD registry. This ensures that all important data resources could be pooled and analysed together in future collaborative studies. Another issue is the heterogeneity of the PCD patient population concerning the diagnostic evaluation; diagnostic results can be normal or very subtle in some patients who almost certainly have PCD. Therefore, we have grouped the patients into three subgroups based on the recent diagnostics guidelines from the ERS Task Force [11] and are planning to take the level of diagnostic certainty into account in all analyses. Participating centres will be encouraged to follow these guidelines for new patients and to re-evaluate the diagnosis of patients in cases where not all testing has been concluded. Although the iPCD Cohort is an important data source for statistically meaningful research, it is not truly representative of the PCD patient population. Countries that have not developed diagnostic facilities yet are under-represented. PCD is underdiagnosed in adults, and most centres follow mainly children and younger adults. Thus, adult patients, especially those of older age, are under-represented in the Cohort.

Most published studies on PCD include a small number of patients. The largest published study on PCD was in the framework of the previous ERS PCD Task Force, which included 1192 patients from 26 countries [7]. In that study the researchers collected basic data from questionnaires addressed to the clinical centres of participating countries. They found a slightly higher proportion of male patients and 85% of the
patients were aged <20 years. Recently, data on 201 patients from the international PCD registry have been described [26]. The authors observed 55% males and almost 50% patients aged <18 years. 22% of the included patients have only a clinical diagnosis compared with 35% in the iPCD Cohort. As these two datasets were both built in the framework of BESTCILIA there is an overlap in the patients they include. The international PCD registry aims to collect prospective data of PCD patients in a standardised way to use for future studies, while the iPCD Cohort combines available retrospective datasets of PCD patients in pooled analyses to answer important questions on PCD in a large study population. Although well-designed prospective studies have obvious strengths, it is of great importance to use all available data resources for research in rare diseases such as PCD, where there is still little evidence and management of patients is primarily based on experience derived from other diseases, such as cystic fibrosis.

The iPCD Cohort is a valuable resource for epidemiological studies in PCD and it can be further enriched and used in the framework of the BEAT-PCD EU COST Action. BEAT-PCD aims to set up a global, Europe-led multidisciplinary network of PCD clinicians and researchers, and the aims of the epidemiological work group include using this dataset for projects on PCD epidemiology. More centres have expressed an interest in contributing data to the cohort. Some groups had originally contributed a basic cross-sectional dataset and are now adding retrospectively collected repeated measurements (longitudinal data) or they are adding new patients to their datasets. Ongoing studies analyse growth, lung function, diagnostic tests, clinical manifestations, neonatal symptoms and the role of lobectomy in PCD, using the iPCD Cohort dataset. The iPCD Cohort also contains data on treatments, imaging and microbiology, which can be used and to identify patients for nested studies on PCD.

This dataset is now available to be further exploited and offers a unique opportunity to study PCD in a large international patient-based cohort with sufficient statistical power. Results will help to improve survival, care and quality of life of PCD patients.

Acknowledgements
We thank all the patients with PCD in the cohort and their families, and especially the PCD patient organisations for their close collaboration. We also thank all the researchers in the participating centres who were involved in data collection and data entry, and worked closely with us through the whole process of participating to the iPCD Cohort. We acknowledge Jingying Wang (Institute of Social and Preventive Medicine, University of Bern, Bern, Switzerland) for her help and technical support in building the REDCap dataset.

Author contributions: C.E. Kuehni developed the concept and designed the study. C.E. Kuehni, M. Goutaki and E. Maurer identified eligible datasets, drafted agreements and prepared the standardised dataset. M. Goutaki and F.S. Halbeisen cleaned and standardised the data, and performed the statistical analyses. All other authors participated in discussions for the development of the study and contributed data. C.E. Kuehni, M. Goutaki, F.S. Halbeisen and J.S. Lucas drafted the manuscript; all authors contributed to iterations and approved the final version. C.E. Kuehni and M. Goutaki take final responsibility for the contents.

https://doi.org/10.1183/13993003.01181-2016
References