





# Pulmonary Vascular Research Institute GoDeep: A meta-registry merging deep phenotyping data from international PH reference centers

Raphael W. Majeed<sup>1,2</sup>  | Martin R. Wilkins<sup>3</sup> | Luke Howard<sup>3</sup> |  
Paul M. Hassoun<sup>4</sup> | Anastasia Anthi<sup>5</sup> | Hector R. Cajigas<sup>6</sup> | John Cannon<sup>7</sup> |  
Stephen Y. Chan<sup>8</sup>  | Victoria Damonte<sup>9</sup> | Jean Elwing<sup>10</sup> | Kai Förster<sup>11</sup> |  
Robert Frantz<sup>12</sup>  | Stefano Ghio<sup>13</sup> | Imad Al Ghouleh<sup>8</sup> | Anne Hilgendorff<sup>11</sup> |  
Arun Jose<sup>10</sup>  | Ernesto Juaneda<sup>9</sup> | David G. Kiely<sup>14</sup> | Allan Lawrie<sup>14</sup>  |  
Stylianios E. Orfanos<sup>5</sup> | Antonella Pepe<sup>13</sup> | Joanna Pepke-Zaba<sup>7</sup> |  
Yuriy Sirenko<sup>15</sup> | Andrew J. Swett<sup>16</sup> | Olena Torbas<sup>15</sup> | Roham T. Zamanian<sup>16</sup> |  
Kurt Marquardt<sup>1</sup> | Achim Michel-Backofen<sup>1</sup> | Tobiah Antoine<sup>1</sup> |  
Jochen Wilhelm<sup>17</sup> | Stephanie Barwick<sup>18</sup> | Phillipp Krieb<sup>1</sup> |  
Meike Fuenderich<sup>1</sup> | Patrick Fischer<sup>1</sup> | Henning Gall<sup>1</sup>  |  
Hossein-Ardeschir Ghofrani<sup>1,17</sup> | Friedrich Grimminger<sup>1,17</sup> | Khodr Tello<sup>1,17</sup> |  
Manuel J. Richter<sup>1,17</sup> | Werner Seeger<sup>1,17</sup> 

<sup>1</sup>Department of Internal Medicine, Universities of Giessen and Marburg Lung Center (UGMLC), German Center for Lung Research (DZL), Giessen, Germany

<sup>2</sup>Institute of Medical Informatics, RWTH Aachen University, Aachen, Germany

<sup>3</sup>National Heart and Lung Institute and Imperial College, London NHS Healthcare Trust, London, UK

<sup>4</sup>Department of Medicine, Division of Pulmonary and Critical Care Medicine, Johns Hopkins University School of Medicine, Baltimore, Maryland, USA

<sup>5</sup>1st Department of Critical Care, National & Kapodistrian University of Athens Medical School and Pulmonary Hypertension Clinic, Evangelismos General Hospital, Athens, Greece

<sup>6</sup>Division of Pulmonary and Critical Care Medicine, Mayo Clinic, Rochester, New York, USA

<sup>7</sup>Pulmonary Vascular Diseases Unit, Royal Papworth Hospital, Cambridge Biomedical Campus, Cambridge, UK

<sup>8</sup>Department of Medicine, Division of Cardiology, Center for Pulmonary Vascular Biology and Medicine, Pittsburgh Heart, Lung, Blood, and Vascular Medicine Institute, University of Pittsburgh School of Medicine and University of Pittsburgh Medical Center, Pittsburgh, Pennsylvania, USA

<sup>9</sup>Hospital de Niños, Hospital Privado Universitario, Clinica Universitaria Reina Fabiola and Instituto Oulton-Catholic, University of Córdoba, Córdoba, Argentina

<sup>10</sup>Division of Pulmonary, Critical Care and Sleep Medicine, University of Cincinnati College of Medicine, Cincinnati, Ohio, USA

<sup>11</sup>Ludwig-Maximilians University of Munich, München, Germany

<sup>12</sup>Department of Cardiology, Mayo Clinic, Rochester, New York, USA

<sup>13</sup>Fondazione IRCCS Policlinico S.Matteo, Pavia, Italy

Manuel J. Richter contributed equally to this work.

This is an open access article under the terms of the Creative Commons Attribution-NonCommercial License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited and is not used for commercial purposes.

© 2022 The Authors. *Pulmonary Circulation* published by John Wiley & Sons Ltd on behalf of Pulmonary Vascular Research Institute.

<sup>14</sup>Sheffield Pulmonary Vascular Disease Unit, Royal Hallamshire Hospital and University of Sheffield, Sheffield, UK

<sup>15</sup>Department of Symptomatic Hypertension, “National Scientific Center ‘The M.D. Strazhesko Institute of Cardiology’” of National Academy of Medical Science, Kyiv, Ukraine

<sup>16</sup>Division of Pulmonary, Allergy, and Critical Care, and Vera Moulton Wall Center for Pulmonary Vascular Disease, Stanford University, Stanford, California, USA

<sup>17</sup>Institute for Lung Health (ILH), Cardio-Pulmonary Institute (CPI), Giessen, Germany

<sup>18</sup>Pulmonary Vascular Research Institute (PVRI), Canterbury, UK

### Correspondence

Werner Seeger, Department of Internal Medicine, Justus Liebig University Giessen, Universities of Giessen and Marburg Lung Center (UGMLC), Klinikstrasse 33, 35392 Giessen, Germany. Email: [werner.seeger@ugmlc.de](mailto:werner.seeger@ugmlc.de)

### Funding information

Pulmonary Vascular Research Institute (PVRI); Cardiovascular Medical Research and Education Fund (CMREF)

### Abstract

The Pulmonary Vascular Research Institute GoDeep meta-registry is a collaboration of pulmonary hypertension (PH) reference centers across the globe. Merging worldwide PH data in a central meta-registry to allow advanced analysis of the heterogeneity of PH and its groups/subgroups on a worldwide geographical, ethnical, and etiological landscape (ClinTrials.gov NCT05329714). Retrospective and prospective PH patient data (diagnosis based on catheterization; individuals with exclusion of PH are included as a comparator group) are mapped to a common clinical parameter set of more than 350 items, anonymized and electronically exported to a central server. Use and access is decided by the GoDeep steering board, where each center has one vote. As of April 2022, GoDeep comprised 15,742 individuals with 1.9 million data points from eight PH centers. Geographic distribution comprises 3990 enrollees (25%) from America and 11,752 (75%) from Europe. Eighty-nine percent were diagnosed with PH and 11% were classified as not PH and provided a comparator group. The retrospective observation period is an average of 3.5 years (standard error of the mean 0.04), with 1159 PH patients followed for over 10 years. Pulmonary arterial hypertension represents the largest PH group (42.6%), followed by Group 2 (21.7%), Group 3 (17.3%), Group 4 (15.2%), and Group 5 (3.3%). The age distribution spans several decades, with patients 60 years or older comprising 60%. The majority of patients met an intermediate risk profile upon diagnosis. Data entry from a further six centers is ongoing, and negotiations with >10 centers worldwide have commenced. Using electronic interface-based automated retrospective and prospective data transfer, GoDeep aims to provide in-depth epidemiological and etiological understanding of PH and its various groups/subgroups on a global scale, offering insights for improved management.

### KEYWORDS

deep phenotyping, meta-registry, outcome, pulmonary hypertension, risk assessment, worldwide outreach

## INTRODUCTION

Pulmonary hypertension (PH), defined by an elevated resting mean pulmonary artery pressure (mPAP), is a chronic condition with multifactorial etiology.<sup>1,2</sup> Patients are classified into five major groups by international consensus.<sup>2</sup> These groups are based on clinical,

pathobiological and hemodynamic considerations and comprise pulmonary arterial hypertension (PAH), Group 1; PH due to left-sided heart disease, Group 2; PH due to lung disease or hypoxia, Group 3; chronic thromboembolic pulmonary hypertension, Group 4; and pulmonary hypertension with unclear or multifactorial mechanisms, Group 5.

Usually considered a rare disease, due to the historical focus on idiopathic PAH (IPAH), current estimates calculate an age-dependent PH prevalence of about 1% of the global population, which increases up to 10% in the elderly.<sup>3</sup> Worldwide, PH Group 2 is thought to affect 5%–10% of elderly individuals (>65 years),<sup>3</sup> and in-depth analyses reveal that PH Group 3 encompasses another very large group of PH, far outnumbering Group 1 patients.<sup>4</sup> Global epidemiological data indicate that approximately 20% of all hospitalized patients might present with some variant of PH or pulmonary pressure elevation.<sup>5</sup> Taken together, these findings support the contention that PH is substantially more prevalent than commonly assumed and constitutes a comorbidity of global significance. Such a view is further supported by the observation that individuals with only mildly elevated pulmonary pressures, regardless of etiology, have reduced survival.<sup>6</sup>

National PH registries, largely based in Europe and the United States, have offered useful insights into PH through advanced clinical phenotyping and risk assessment leading to treatment guidance.<sup>7</sup> Their main limitation is that they are parochial and speak mostly to their own population. Therefore, the Pulmonary Vascular Research Institute (PVRI) decided to take advantage of its representation on all continents to set up a new global meta-registry, named PVRI GoDeep, aiming to collect and amalgamate deep phenotype data contained in various local registries throughout the world. Based on user-friendly electronic interfacing of anonymized data sets and using a commonly agreed extensive phenotype parameter list, it collects both retrospective and prospective data. The ambition is to capture in-depth clinical phenotypes already available and progressively collected in the various contributing centers for cluster-analysis of PH patients and trend analysis,<sup>8–10</sup> with plans to expand to include sophisticated new imaging techniques,<sup>11</sup> omics analyzes<sup>12</sup> and data from wearable devices.<sup>13</sup> We expect that such a comprehensive worldwide state-of-the-art registry will impact our understanding of PH as a global condition, expand our knowledge of the influence of regional etiological factors, and enhance our ability to guide PH treatment against this background. This report describes the basic characteristics of GoDeep, based on data sets from eight renowned PH centers, with several further being in the status of data transfer preparation or contract completion.

The main goals of GoDeep can be summarized as follows: establishment of the largest international collaborative PH registry; framework for world-wide comprehensive research; establishment of a deep phenotyping

PH data bank spanning over all continents and development of GoDeep into the go-to place for PH specialists and companies addressing pulmonary vascular diseases around the world.

## METHODS

### Basic setup

As a meta-registry, PVRI GoDeep includes patients from existing PH cohorts with both retrospective data entry and further prospective data collection. The main inclusion criterion is PH defined by elevated pulmonary artery pressure as diagnosed by right heart catheterization (RHC), according to the definition of the PH World Symposium at the time of diagnosis.<sup>2</sup> Individuals in which manifest PH was excluded by right heart catheterization, provide a comparator group.

Integrating retrospective data from separate and independent cohorts at different centers across multiple countries/continents requires linking data points to a common set of parameters. The common parameter list was defined in three steps. An initial meeting between two founding PH centers, namely Imperial College London, United Kingdom, and Justus-Liebig University Giessen, Germany, provided a draft version of relevant parameters using expert consensus. This initial list of parameters was then extended by reconciliation with the PH center at Johns-Hopkins University (Baltimore, USA). In the third step, the list was merged with the parameter list from the International Consortium for Genetic Studies in PAH (PAH-ICON) and further discussed and rated by six domain experts, resulting in a total of more than 350 items. These parameters fall into four different categories (mandatory, essential, recommended, extended), and risk assessment tools are integrated and currently comprise the 4-strata risk tool<sup>14</sup> and the 3-strata European PH risk stratification.<sup>15</sup> The following items will be included: Mandatory: Date of birth, sex, diagnostic classification, right heart catheter defining onset of disease; Essential: Ethnicity, survival status, date/age at diagnosis, WHO functional class, 6-min-walking test, pulmonary function test, medication; Recommended: Onset of symptoms, comorbidities, echocardiography, electrocardiogram, blood gas analysis, laboratory; Extended: Spirometry, quality of life, cardiac magnetic resonance imaging, biomaterials, genetic analysis. A separate publication detailing this parameter list is currently in preparation.

For interoperability, all items were mapped to the international standard terminologies Systematized Nomenclature of Medicine—Clinical Terms (SNOMED-CT) and Logical Observation Identifiers Names and Codes (LOINC). We applied for new LOINC codes in

seven cases where no international standard codes were available.

In addition to the three founding PH centers, five additional centers have joined and connected their data to PVRI GoDeep (as of April 2022): Stanford (USA), Pittsburgh (USA); Cordoba (Argentina), Sheffield (UK), Pavia (Italy). Eight PH registries thus contributed data to the analysis presented in this publication. An agreement is in place with six further centers and their data entry is currently in preparation: Munich (Germany), Cambridge (UK), Rochester (USA), Cincinnati (USA), Athens (Greece), Kiev (Ukraine).

## Use and access rules

The role of contributing PH centers in PVRI GoDeep is twofold: Each site commits to providing its full set of retrospective data collection as well as further prospective data (matching the common variable list as close as possible) to be updated on regular intervals (see section “data quality and descriptive analysis” below). In turn, each center is granted access to all data in the meta-registry. Moreover, each site has one seat in the PVRI GoDeep steering-board that decides on all requests of scientific analysis of the GoDeep data and discusses future directions and developments for PVRI GoDeep. Access to data is provided on three levels. The first (public) level includes detailed metadata in the form of parameter lists, data entry forms for REDCap, aggregated data, summaries, and showcases accessible to anyone via the PVRI GoDeep website. The second level requires user login and enables access to custom real-time feasibility analyzes via the i2b2 query tool,<sup>16</sup> returning patient counts and simple breakdown statistics. The third level of access includes detailed statistical analyzes and access to selected raw data as well as support by GoDeep statisticians to answer specific research questions, but requires individual approval by the GoDeep steering board. Such requests for scientific evaluation will primarily originate from members of the GoDeep consortium. In addition, external academic institutions and industrial partners can request data analyzes (not data export) via the steering committee, which will decide on a case-by-case basis. Each contributing GoDeep center is free to decide whether their data package will also be entered into analyzes for external requests.

## Regulatory affairs and data protection

To simplify international regulatory affairs and data protection legislation, our registry accepts only

anonymized data. Anonymization is performed by removing exact date and time information and performing k-anonymization on the remaining quasi-identifiers sex, ethnicity, and birth decade. Instead of absolute date information, all dates are converted to a relative number of months from diagnosis to each visit or death. After this calculation, the diagnosis date for all patients is converted to decade resolution (e.g., 1990s, 2000s). All of these de-identification routines are performed locally at each contributing site. For three sites that have currently entered data into GoDeep, we provided customized anonymization algorithms in the form of R project scripts. Five sites performed the anonymization themselves.

To join PVRI GoDeep, the required formal steps for sites with existing PH cohorts depend on local and national regulations: Typically, local approval by ethics and data protection is required. In the United States, this is handled by the local institutional review board. In the United Kingdom and European Union, separate legal entities are responsible for ethical review (ethical review board) and data protection approval (Caldicott Guardian in UK). Additionally, most sites require that the agreement to participate in GoDeep is based on a detailed data use agreement (DUA).

The University of Giessen/University Hospital Ethics Committee and the responsible data protection officer have approved the PVRI-GoDeep central data repository. Adherence to current data safety recommendations and ethical committee considerations includes the following components: Patient Data Privacy Concept encompassing detailed description of technical and organizational measures for data protection and information security; Metadata Repository for harmonization and annotation of patient-related data across different biobanks and registries/cohorts/studies.

In addition to gaining approval of the central PVRI-GoDeep data repository by the University of Giessen Ethics Committee, each contributing center obtained permission by the local Ethics Committee and the responsible data protection authorities to enter their data into the GoDeep meta-registry in an anonymized fashion. Based on the already available patients' consent for data collection within the local registries, forwarding anonymized data to the meta-registry did not demand further additional consent by the patients, as decided by the respective local authorities.

To simplify these formal processes for interested and contributing sites, we provide documentation and templates for ethical review, data protection, and DUA contracts. With the provided documents, all contributing sites passed local ethical and data protection review easily.



## Standardization and interoperability

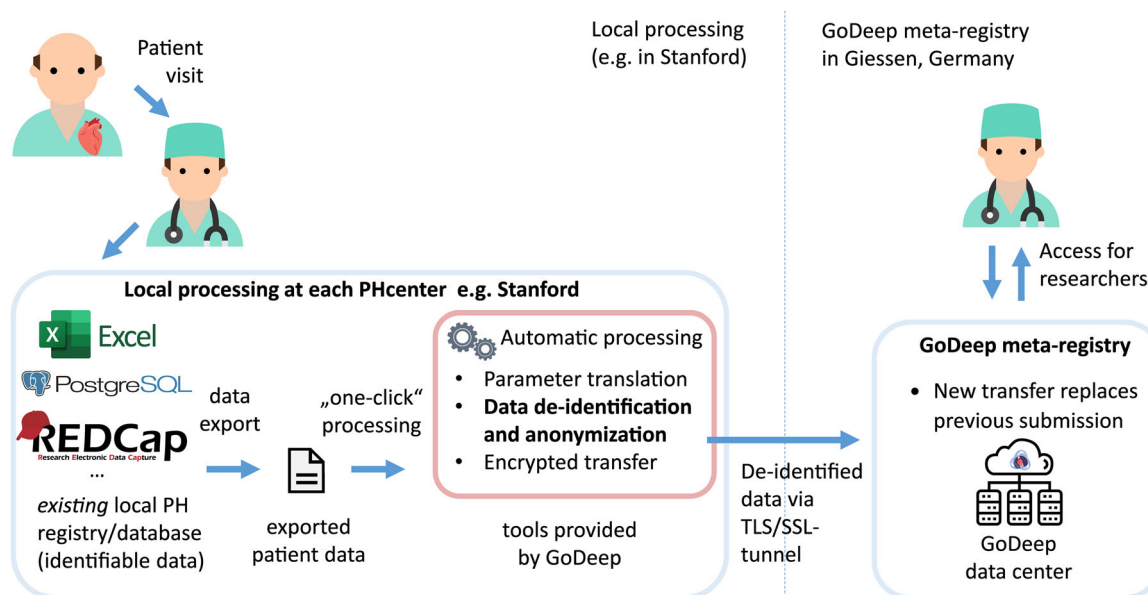
At the central PVRI GoDeep hub, all anonymized data is transformed into standard data structures and standard terminologies to increase interoperability (Figure 1). For structural interoperability, we used the international standard Fast Healthcare Interoperability Resources (FHIR) by Health Layer Seven International (HL7). Primary data structure for persistence are FHIR Bundle collections. Data from each participating site/registry are transformed and compiled into separate Bundle resources. Within each collection, all data points are represented in multiple FHIR Resources from the type of patient, encounter and observation using FHIR R4 standard profiles. Before data from a site/registry is connected to PVRI GoDeep, we annotate local parameters with the standard terminologies SNOMED-CT and LOINC. Through this annotation, a link between local parameters and our GoDeep common data set is established while also ensuring that all FHIR resources use only standard terminology codes.

## Data quality and descriptive statistics

Data from each hospital is updated on a regular basis, usually quarterly. To enable additional visits and corrections in longitudinal data in the anonymized data set, each center always transfers a full data set, which in turn replaces the previous full data set.

For each data transfer, automated data quality analysis is performed: implausible data are identified using time relations (e.g., medication after death, etc.) and boundary checks for values, which were defined by domain experts. Scripts for automated detection of implausible values were programmed by GoDeep biostatisticians using the software “R” ([www.R-project.org](http://www.R-project.org)). These scripts are run centrally when updated data are provided by a PH center and the resulting feedback is given to the corresponding site. With the provided data quality feedback, sites can choose to fix issues locally and resubmit corrected data. Before analysis, implausible data are removed using the same scripts and subsequently treated as missing data.

For breakdown by PH groups, the definitions from the PH World Symposium were used.<sup>17</sup> The first confirmation of PH based on RHC was used as time point of diagnosis of PH. PH classification and diagnosis criteria are regularly updated according to current guidelines and recommendations<sup>2</sup> at the respective centers. Phenotypic drift is addressed by allowing multiple PH group assignments per patient at different points in time. For the descriptive analysis presented in this publication, risk strata were calculated at the earliest point in time per patient. Similarly, the earliest WHO functional class assessment after diagnosis was used. A single patient may fall into multiple PH groups, which may also change over time, for example, after additional diagnostic steps. In these cases, only the latest diagnostic



**FIGURE 1** Data flow from local PH registry to PVRI GoDeep. PH, pulmonary hypertension; PVRI, Pulmonary Vascular Research Institute.

classification is used. If the latest classification includes multiple groups, the patient will count toward all groups assigned at the latest point in time. Time-related biases will be minimized due to periodical updates of survival information, identification of missing survival data, rigorous definition of patient entry, time-points of relevant treatment or diagnosis and corresponding follow-up periods as well as employing time-dependent statistical analyzes.

The descriptive analysis presented in this paper used the statistical software R project (version 4.1.2).

## RESULTS

As of April 2022, PVRI GoDeep contains data on 15,742 enrollees, of which 9884 (63%) are female and 5849 (37%) are male (7 specified other gender). By continent, 3990 subjects (25%) are from the Americas and 11,752 (75%) from Europe. The proportion of female enrollees is higher in the Americas with 71% compared to 60% in Europe. All enrollees underwent right heart catheterization. Those patients, in whom the catheterization excluded manifest PH, were entered into the comparator group ( $n = 1735$ ).

This collective was named “Comparator” group (Table 1). The remaining group with proven PH thus includes 14,007 patients. The majority of these PH

patients have been assigned to Group 1 (42.6%), followed by Group 2 (21.7%), Group 3 (17.3%), Group 4 (15.2%), and Group 5 (3.3%). In total, 447 PH patients were not yet assigned to one of the five groups at the time of data entry into GoDeep as of April 2022, and in total 218 patients (<2%) had more than one group assigned. Concerning the subgroups of PAH (Group 1), Subgroup 1.1 and Subgroup 1.4 clearly dominate both in Europe and the Americas (Table 2). Subgroup 1.3 (drug-induced PH) is much higher in the American population (22% of Group 1) than in the European population (1% of Group 1). The age distribution at the time of diagnosis (first confirmation of PH by RHC) spans over decades, with patients 60 years or older comprising 60% (Figure 2 and Table 1). The average age at PH diagnosis in Europe is 62 years, as compared to 56 years in the Americas. The age distribution of the

**TABLE 2** Distribution of PAH Group1 subgroups by continent

Continent	Group 1.1	Group 1.2	Group 1.3	Group 1.4
America	376 (27%)	21 (2%)	313 (22%)	683 (49%)
Europe	1594 (42%)	133 (4%)	32 (1%)	2032 (54%)
Total	1970 (38%)	155 (3%)	345 (7%)	2715 (52%)

Abbreviation: PAH, pulmonary arterial hypertension.

**TABLE 1** Basic descriptive statistics by PH group and continent

PH group <sup>2</sup>	Group 1	Group 2	Group 3	Group 4	Group 5	Comparator group
Patient count by continent						
America	1935 (58%)	509 (15%)	538 (16%)	184 (5%)	185 (6%)	506
Europe	3948 (38%)	2491 (24%)	1849 (18%)	1913 (18%)	272 (3%)	1229
Total	5883 (43%)	3000 (22%)	2387 (17%)	2097 (15%)	457 (3%)	1735
Age at diagnosis						
0–9	40	2	0	0	0	0
10–19	130	1	7	5	1	3
20–29	361	11	18	47	11	57
30–39	616	46	60	123	32	117
40–49	967	128	197	244	72	235
50–59	1120	353	435	364	123	363
60–69	1297	731	785	535	120	446
70–79	1077	1238	732	565	77	421
80–89	272	477	151	211	21	92
90–99	3	13	2	3	0	1

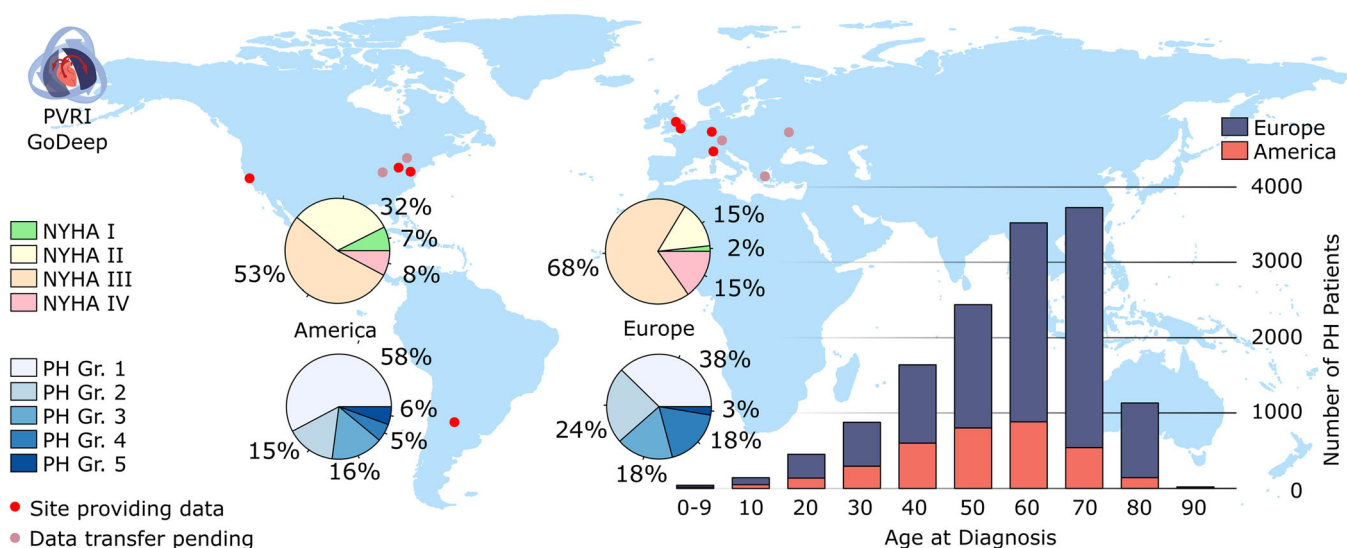
Note: 218 patients (1.6%) had more than one PH group assigned, and 447 patients (3.2%) were not yet definitively assigned to any group.

Abbreviation: PH, pulmonary hypertension.

Comparator group largely corresponds to that of the PH patients.

The majority of patients were assigned to a functional class (NYHA) III (overall 64%) (Figure 2). Comparing Europe with the Americas, the percentage of patients presenting with severe PH at the time of diagnosis (NYHA III and IV) was higher in Europe (85%) than in the Americas (59%). Concerning the risk assessment of the entire population, the majority of PH patients met the criteria of an intermediate risk profile when applying a 3-strata classification<sup>14</sup>

(Table 3). Interestingly, this was true for all groups of PH patients, ranging from 57% in Group 2 to 69% in Group 4. Correspondingly, when applying a 4-strata risk assessment,<sup>14</sup> 61% of the patients were classified as intermediate low or intermediate-high, with the proportion classified as intermediate-high risk more than double that of the intermediate-low risk (45% vs. 16%). Again, this distribution was largely comparable between the different PH groups. When comparing Europe to the Americas, patients in Europe were classified into higher risk groups, with almost double



**FIGURE 2** Distribution and global patterns of PH patients currently entered into the GoDeep meta-registry. NYHA, New York Heart Association functional class; PH, pulmonary hypertension; PH Gr, groups of PH diagnosis.

**TABLE 3** Risk assessment of PH groups

PH group <sup>2</sup>	Group 1	Group 2	Group 3	Group 4	Group 5
Risk assessment according to <sup>15</sup>					
Low risk	1676 (28%)	614 (20%)	427 (18%)	452 (22%)	87 (19%)
Intermediate risk	3796 (65%)	1720 (57%)	1433 (60%)	1439 (69%)	301 (66%)
High risk	281 (5%)	242 (8%)	278 (12%)	96 (5%)	27 (6%)
(Not available)	130 (2%)	424 (14%)	249 (10%)	110 (5%)	42 (9%)
4-strata risk assessment <sup>14</sup>					
Low Risk	685 (12%)	292 (10%)	129 (5%)	195 (9%)	38 (8%)
Intermediate-low risk	1087 (18%)	430 (14%)	353 (15%)	261 (12%)	83 (18%)
Intermediate-high risk	2644 (45%)	1324 (44%)	1006 (42%)	1127 (54%)	198 (43%)
High risk	1096 (19%)	496 (17%)	619 (26%)	393 (19%)	87 (19%)
(Not available)	371 (6%)	457 (15%)	280 (12%)	121 (6%)	51 (11%)

Note: For risk assessment, earliest data points were used. Two hundred and seventeen patients (1%) had more than one PH group assigned, thus the row-sums slightly exceed the total number of patients.

Abbreviation: PH, pulmonary hypertension.

the share of high-risk patients for both risk classifications (8% in European Union vs. 4% in the Americas via 3-strata and 22% vs. 12% via 4-strata) (Table 4).

The mean observation period of all PH patients entered into the GoDeep meta-registry from the various local registries was 3.5 years (standard error of the mean [SEM] 0.04), ranging from <1 to >25 years

**TABLE 4** Descriptive statistics of PH by continent and overall

	America	Europe	Overall
Total	3484	10,523	14,007
Female	2416 (71%)	6273 (60%)	8734
Male	1017 (29%)	4248 (40%)	5265
Other gender	6 (0%)	2 (0%)	8 (0%)
Observation period			
0–2 years	1239 (40%)	5955 (57%)	7194 (53%)
2–10 years	1458 (47%)	3726 (36%)	5184 (38%)
>10 years	417 (13%)	742 (7%)	1159 (9%)
Risk assessment <sup>15</sup>			
Low risk	982 (28%)	2377 (23%)	3359 (24%)
Intermediate risk	2256 (65%)	6464 (61%)	8720 (62%)
High risk	128 (4%)	804 (8%)	933 (7%)
(Not available)	117 (3%)	878 (8%)	995 (7%)
4-strata risk assessment <sup>14</sup>			
Low risk	445 (13%)	927 (9%)	1372 (10%)
Intermediate-low risk	897 (26%)	1287 (12%)	2184 (16%)
Intermediate-high risk	1259 (36%)	5096 (48%)	6355 (45%)
High risk	431 (12%)	2276 (22%)	2707 (19%)
(Not available)	452 (12%)	937 (9%)	1389 (10%)

Note: Percentages are calculated vertically.

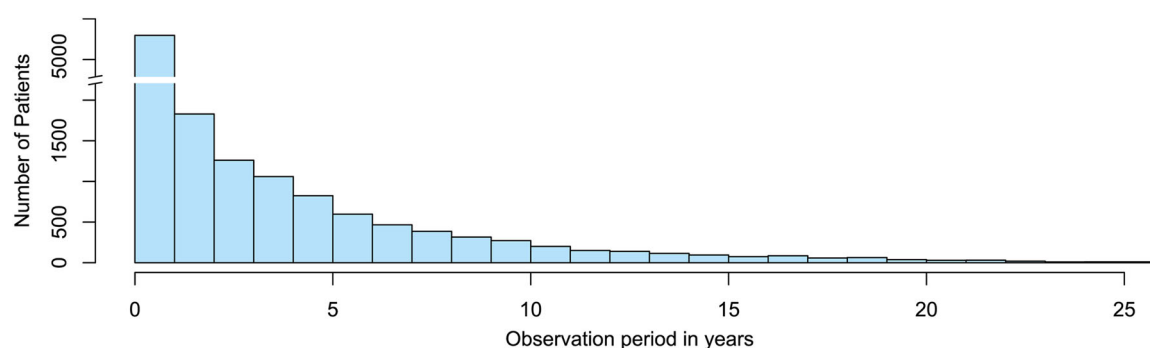
Abbreviation: PH, pulmonary hypertension.

(Figure 3). In total 1159 PH patients were followed for more than 10 years, 742 of these in Europe and 417 of these in the Americas.

## DISCUSSION

Over the past decade, national PH registries have provided an important basis for clinical research and a fundamental understanding of risk stratification.<sup>7</sup> In addition, this has substantially influenced clinical trial design.<sup>7</sup> However, the PH landscape is changing, and better understanding requires larger databases that inform on a global scale, enabling the development of new assessment tools. The PVRI GoDeep meta-registry aims to address this challenge by integrating data from multiple PH reference centers worldwide. With eight centers, it already comprises the largest deep-phenotyped PH databank worldwide (>15,000 individuals). With a further six sites in the process of integrating their PH cohorts, as well as negotiations currently going on with >10 additional centers worldwide, enrollment of more than 30,000 individuals is to be expected when fully established.

In Europe and the Americas, Group 1 still represents the largest group (42.6%), with the predominant Subgroups 1.1 (idiopathic PAH) and 1.4 (associated PAH). The strikingly higher percentage of drug-induced PH within the American as compared to the European Group 1 patients (22% vs. 1%) requires further in-depth analysis. Moreover, there is an increasing representation of PH patients classified as Group 2, 3, and 4 in the PH registries in both continents. This is not surprising, as quantitatively the PH Groups 2 and 3 are far the largest ones worldwide. PH-specific therapies originally designed for PAH patients have now also been shown to be beneficial in Group 3<sup>17</sup> and Group 4,<sup>18,19</sup> and a multitude of studies currently focus on these patient populations. In this regard, PVRI GoDeep offers a



**FIGURE 3** Retrospective observation periods of PH patients collected in GoDeep as of April 2022. PH, pulmonary hypertension.



comprehensive database for further “insilico” exploration of specific PH treatment concepts across various PH groups/subgroups, including analysis of risk stratification,<sup>18</sup> toleration/adherence and outcome.<sup>19</sup> Group 5 may become more important as the focus shifts to PH worldwide. We decided to include clinical and hemodynamic data of the so-called comparator group, individuals, in whom clinical and non-invasive testing led to suspicion of PH, but which was not confirmed by the RHC results. Fifty-one percent of these enrollees have mean pulmonary artery pressures between 20 and 24 mmHg (with pulmonary vascular resistance values <3 WU) when measured under baseline conditions. This subgroup might reveal PH upon exercising or fluid challenge, which was not systematically investigated in this cohort. Cluster analysis of this putative “borderline” group may well become of interest for future preventive types of studies.

Comparing the patients entered from the American versus the European centers showed higher female percentage and younger mean age of the former population, whereas the assignment of PH patients to the various groups and subgroups was rather similar between the Americas and Europe, except for drug-induced PH (Subgroup 1.3). Interestingly, the percentage of severe and very severe PH was higher in Europe as compared to America: 85% of all PH patients presented with NYHA III and IV upon first assessment of this variable, as compared to 59% in the American reference centers. This is also reflected by the 3-strata and 4-strata risk classification. Whether this difference is due to higher “awareness” of PH in the Americas, resulting in earlier transmission of the patients to reference centers, will demand more in-depth analysis. Of note, the risk classification is not yet validated for all PH groups, albeit there is emerging data for Group 4.<sup>20</sup> Further validation of the risk classification using the GoDeep registry is mandatory to perform.

It is of particular interest that from both the American and the European PH reference centers, long-term observation data were entered into the meta-registry. To the best of our knowledge, this is the only registry worldwide to include >1000 patients with detailed follow-up data over more than 10 years, some of those extending to >25 years. Detailed analysis of these patients may allow identifying individual characteristics and/or therapeutic measures, which are linked with “longevity with PH.”

As of April 2022, GoDeep includes the data sets collected in the various PH reference centers over the past years/decades, with individual patient-level data entry being strictly based on the verification of PH diagnosis by right heart catheterization. The only

exception in this respect is the setting up of a neonatal/early pediatric PH cohort, for which a specifically designed parameter list is currently being discussed. This “retrospective approach” also includes a multitude of variables in addition to the RHC-based hemodynamic data, collected by but also partially differing among the various PH centers. For the further prospective data collection, GoDeep has now defined a common list of parameters falling into different categories (mandatory, essential, recommended, extended), integrating also 3-strata and 4-strata risk assessment tools, to be published in detail separately. This parameter set will allow reliable assessment of the various PH groups and subgroups and permit further deeper phenotyping, cluster analysis and assessment of treatment concepts and clinical follow-up. To largely avoid missing data, GoDeep performs periodical electronic interface-based automated data update from the local registries.

PVRI GoDeep is run under the patronage of the PVRI (<https://pvriinstitute.org>) and operated under the auspices of the University of Giessen/Giessen PH center, Germany. A fair balance of interests between the owners of the local registries and PVRI GoDeep is given by the authority of the common steering board, in which each contributing center has one vote, with clearly defined use and access rules. A competence team has been set up, serving all participating registries for data harmonization, automated data transfer and state of the art bioinformatic/statistical analysis. AI competence will be additionally integrated into the analysis team. As for quality control, this is primarily based on the high standards of the contributing PH reference centers. In addition to the respective local quality control measures, a quality control process has been established on the central GoDeep meta-registry level, with detailed feedback on implausible or missing data going to the local centers for correction.

Registry data represent the cornerstone for outcome prediction and development of risk stratification tools, as demonstrated by the French, Swedish, Comparative, Prospective Registry of Newly Initiated Therapies for Pulmonary Hypertension, and Registry to Evaluate Early and Long-Term Pulmonary Arterial Hypertension Disease Management registries.<sup>14,15,21,22</sup> PVRI GoDeep incorporates all currently available risk strata and will allow validation in all PH groups/subgroups, with potential further development leveraging from specific items of the common data trunk as well as from advanced modules, including cross-validation in differently defined cohorts within the meta-registry due to its large database. In addition, it is possible to explore the impact of different hemodynamic definitions among various PH groups. The global approach of PVRI GoDeep

allows a better understanding of differences in PH classification among various ethnicities or regions/countries/continents. This may have implications for the current clinical classification. Regional and international temporal trends referring to PH burden, health care system, environmental factors, demographics, disease severity, and use or availability of specific PH medication can also be explored. Longitudinal data will enable advanced assessment of regional differences and trends in disease progression stratified by background therapies or up-front strategy. Childhood PH and early adult-onset PH can be compared to late adult-onset PH, not possible with most existing registries.

It was a major challenge of GoDeep to overcome the differences in the collection of phenotypic data at an individual patient level among the international contributors. To this end, a detailed parameter list with precise definition for each variable has been set up. Data protection via anonymization causes some limitation of data analysis, such as for example, concerning the exact date of birth (substituted by a birth period), but deep granularity of the vast majority of variables is nevertheless given, providing a rich source for future scientific evaluation. It is a further limitation of the current GoDeep status that the centers having entered their data are largely located in Europe and the United States. However, negotiations and initial steps of processing have been started with centers all around the world. In several cases, the GoDeep logistics team currently supports local activities to convert a given registry into a format, which will then allow establishing an electronic interface with the GoDeep meta-registry. It is the goal of GoDeep to integrate PH reference centers from all continents within the next few years.

## CONCLUSION

Constructed as a global collaborative meta-registry of PH reference centers, PVRI GoDeep already contains the largest phenotypic PH patient data set currently available. This meta-registry will provide the framework for future comprehensive research in the entire field of PH, including its different groups and sub-groups. Combining deep phenotyping with worldwide outreach, PVRI GoDeep aims to offer insights into specific geographical and ethnical profiles of PH, to deepen the epidemiological, clinical and molecular understanding of this disease, to provide information on and comparison with “built-in” subcohorts and “in-silico control groups,” and to promote strategies for improved individualization of PH treatment.

## AUTHOR CONTRIBUTIONS

Raphael W. Majeed, Martin R. Wilkins, Paul M. Hassoun, Henning Gall, Hossein-Ardeschir Ghofrani, Friedrich Grimminger, Khodr Tello, Manuel J. Richter, and Werner Seeger contributed to the conception and design of the work and interpretation of data. Luke Howard, Anastasia Anthi, Hector R. Cajigas, John Cannon, Stephen Y. Chan, Victoria Damonte, Jean Elwing, Kai Förster, Robert Frantz, Stefano Ghio, Imad Al Ghouleh, Anne Hilgendorff, Arun Jose, Ernesto Juaneda, David G Kiely, Allan Lawrie, Stylianos E. Orfanos, Antonella Pepe, Joanna Pepke-Zaba, Yuriy Sirenko, Andrew J. Swett, Olena Torbas, and Roham T. Zamanian contributed to the acquisition of data. Kurt Marquardt, Achim Michel-Backofen, Tobiah Antoine, Jochen Wilhelm, Stephanie Barwick, Phillipp Krieb, Meike Fuenderich, and Patrick Fischer contributed to the analysis of data. Raphael W. Majeed, Manuel J. Richter, and Werner Seeger have also drafted the work and contributed to editing the manuscript.

## ACKNOWLEDGMENTS

We thank the following members from participating institutions for assisting in the preparation of data transfers: John Wharton (London), Drew Hsi (Stanford), Adam Handen (Pittsburgh), Chindu John (Cambridge), Hao Wang (Rochester), Brett Harnett (Cincinnati), Friederike Häfner (Munich), Benjamin Schubert (Munich), Steven Wood (Sheffield). This study was funded by the Pulmonary Vascular Research Institute (PVRI) and the Cardiovascular Medical Research and Education Fund (CMREF). Open Access funding enabled and organized by Projekt DEAL.

## CONFLICTS OF INTEREST

AL has received support and fees from Janssen, Bayer, Novartis, GSK, Alexion, and IQVIA Ltd. DK has received personal fees for consultancy work giving education talks and participation in steering committees from Acceleron, Ferrer, GSK, and Janssen. His department has received grants for research from GSK and Janssen. EJ is GSL Investigator. HAG has received fees from Actelion, AstraZeneca, Bayer, GSK, Janssen-Cilag, Lilly, Novartis, OMT, Pfizer, and United Therapeutics. HG has received fees from Actelion, AstraZeneca, Bayer, GSK, Janssen-Cilag, Lilly, Novartis, OMT, Pfizer, and United Therapeutics. IAG has held research grants from the United Therapeutics and Gilead Sciences Research Scholars Programs. JC has served as a consultant for Acceleron Pharma, Actelion, GSK, and Merck and has received speaker fees from Actelion and GSK. JME has served as a consultant for United Therapeutics, Altavant, Aerovate, Bayer, Gossamer

Bio, and Liquida. She has participated in clinical research funded by Janssen, United Therapeutics, Liquidia, Phase Bio, Gossamer Bio, Bayer, Acceleron, Altavant, and Aerovate. KT has received speaker fees from Janssen. MJR has received support from Janssen Pharmaceutica and Bayer Pharma AG, as well as speaker fees from Janssen Pharmaceutica and OMT. Stephen Y. Chan has served as a consultant for Acceleron Pharma and United Therapeutics; SYC (Stephen Y. Chan) has held research grants from Actelion, Bayer, and Pfizer. SYC has filed patent applications regarding metabolic targeting in pulmonary hypertension. SYC is a director, officer, and shareholder of Synhale Therapeutics. WS has received consultancy fees from Abivax, Lung Biotechnology, Liquidia, Medspray, Pieris, United Therapeutics, and Vectura. The remaining authors declare no conflict of interest.

## ETHICS STATEMENT

The investigation conforms with the principles outlined in the Declaration of Helsinki and was approved by the ethics committee of the Faculty of Medicine at the University of Giessen (approval number 30/19).

## ORCID

Raphael W. Majeed  <http://orcid.org/0000-0002-8260-8815>

Stephen Y. Chan  <http://orcid.org/0000-0002-9520-7527>

Robert Frantz  <http://orcid.org/0000-0003-4128-3978>

Arun Jose  <http://orcid.org/0000-0001-7109-2692>

Allan Lawrie  <http://orcid.org/0000-0003-4192-9505>

Henning Gall  <http://orcid.org/0000-0001-7016-7373>

Werner Seeger  <https://orcid.org/0000-0003-1946-0894>

## REFERENCES

1. Naeije R, Richter MJ, Rubin LJ. The physiologic basis of pulmonary arterial hypertension. *Eur Respir J*. 2022;59(6):2102334.
2. Simonneau G, Montani D, Celermajer DS, Denton CP, Gatzoulis MA, Krowka M, Williams PG, Souza R. Haemodynamic definitions and updated clinical classification of pulmonary hypertension. *Eur Respir J*. 2019;53(1):1801913.
3. Hoeper MM, Humbert M, Souza R, Idrees M, Kawut SM, Sliwa-Hahnle K, Jing ZC, Gibbs JS. A global view of pulmonary hypertension. *Lancet Respir Med*. 2016;4(4):306–22.
4. Gillmeyer KR, Miller DR, Glickman ME, Qian SX, Klings ES, Maron BA, Hanlon JT, Rinne ST, Wiener RS. Outcomes of pulmonary vasodilator use in Veterans with pulmonary hypertension associated with left heart disease and lung disease. *Pulm Circ*. 2021;11(2):20458940211001714.
5. Sliwa K, Wilkinson D, Hansen C, Ntyintyane L, Tibazarwa K, Becker A, Stewart S. Spectrum of heart disease and risk factors in a black urban population in South Africa (the Heart of Soweto Study): a cohort study. *Lancet*. 2008;371(9616):915–22.
6. Stewart S, Chan YK, Playford D, Strange GA. Mild pulmonary hypertension and premature mortality among 154 956 men and women undergoing routine echocardiography. *Eur Respir J*. 2021;59(1):2100832.
7. McGoon MD, Benza RL, Escobedo-Subias P, Jiang X, Miller DP, Peacock AJ, Pepke-Zaba J, Pulido T, Rich S, Rosenkranz S, Suissa S, Humbert M. Pulmonary arterial hypertension: epidemiology and registries. *J Am Coll Cardiol*. 2013;62(25 Suppl):D51–9.
8. Hoeper MM, Pausch C, Grünig E, Klose H, Staehler G, Huscher D, Pittrow D, Olsson KM, Vizza CD, Gall H, Benjamin N, Distler O, Opitz C, Gibbs J, Delcroix M, Ghofrani HA, Rosenkranz S, Ewert R, Kaemmerer H, Lange TJ, Kabitz HJ, Skowasch D, Skride A, Jureviciene E, Paleviciute E, Miliauskas S, Claussen M, Behr J, Milger K, Halank M, Wilkens H, Wirtz H, Pfeuffer-Jovic E, Harbaum L, Scholtz W, Dumitrescu D, Bruch L, Coghlan G, Neurohr C, Tsangaris I, Gorenflo M, Scelsi L, Vonk-Noordegraaf A, Ulrich S, Held M. Idiopathic pulmonary arterial hypertension phenotypes determined by cluster analysis from the COMPERA registry. *J Heart Lung Transplant*. 2020;39(12):1435–44.
9. Vizza CD, Hoeper MM, Huscher D, Pittrow D, Benjamin N, Olsson KM, Ghofrani HA, Held M, Klose H, Lange T, Rosenkranz S, Dumitrescu D, Badagliacca R, Claussen M, Halank M, Vonk-Noordegraaf A, Skowasch D, Ewert R, Gibbs J, Delcroix M, Skride A, Coghlan G, Ulrich S, Opitz C, Kaemmerer H, Distler O, Grünig E. Pulmonary hypertension in patients with COPD: results from the Comparative, Prospective Registry of Newly Initiated Therapies for Pulmonary Hypertension (COMPERA). *Chest*. 2021;160(2):678–89.
10. Hoeper MM, Pausch C, Grünig E, Staehler G, Huscher D, Pittrow D, Olsson KM, Vizza CD, Gall H, Distler O, Opitz C, Gibbs JSR, Delcroix M, Ghofrani HA, Rosenkranz S, Park DH, Ewert R, Kaemmerer H, Lange TJ, Kabitz HJ, Skowasch D, Skride A, Claussen M, Behr J, Milger K, Halank M, Wilkens H, Seyfarth HJ, Held M, Dumitrescu D, Tsangaris I, Vonk-Noordegraaf A, Ulrich S, Klose H. Temporal trends in pulmonary arterial hypertension: results from the COMPERA registry. *Eur Respir J*. 2021;59(6):2102024.
11. Lewis RA, Johns CS, Coglian M, Capener D, Tubman E, Elliot CA, Charalampopoulos A, Sabroe I, Thompson AAR, Billings CG, Hamilton N, Baster K, Laud PJ, Hickey PM, Middleton J, Armstrong IJ, Hurdman JA, Lawrie A, Rothman AMK, Wild JM, Condliffe R, Swift AJ, Kiely DG. Identification of cardiac magnetic resonance imaging thresholds for risk stratification in pulmonary arterial hypertension. *Am J Respir Crit Care Med*. 2020;201(4):458–68.
12. Rhodes CJ, Wharton J, Ghataorhe P, Watson G, Girerd B, Howard LS, Gibbs JSR, Condliffe R, Elliot CA, Kiely DG, Simonneau G, Montani D, Sitbon O, Gall H, Schermuly RT, Ghofrani HA, Lawrie A, Humbert M, Wilkins MR. Plasma proteome analysis in patients with pulmonary arterial hypertension: an observational cohort study. *Lancet Respir Med*. 2017;4(9):717–26.
13. Wesley Milks M, Sahay S, Benza RL, Farber HW. Risk assessment in patients with pulmonary arterial hypertension in the era of COVID 19 pandemic and the telehealth

- revolution: state of the art review. *J Heart Lung Transplant*. 2020;40(3):172–82.
14. Hoepfer MM, Pausch C, Olsson KM, Huscher D, Pittrow D, Grünig E, Staehler G, Vizza CD, Gall H, Distler O, Opitz C, Gibbs JSR, Delcroix M, Ghofrani HA, Park DH, Ewert R, Kaemmerer H, Kabitz HJ, Skowasch D, Behr J, Milger K, Halank M, Wilkens H, Seyfarth HJ, Held M, Dumitrescu D, Tsangaris I, Vonk-Noordegraaf A, Ulrich S, Klose H, Claussen M, Lange TJ, Rosenkranz S. COMPERA 2.0: a refined 4-strata risk assessment model for pulmonary arterial hypertension. *Eur Respir J*. 2022;60(1):2102311.
  15. Kylhammar D, Kjellström B, Hjalmarsson C, Jansson K, Nisell M, Söderberg S, Wikström G, Rådegran G. A comprehensive risk stratification at early follow-up determines prognosis in pulmonary arterial hypertension. *Eur Heart J*. 2017;39(47):4175–81.
  16. Murphy SN, Weber G, Mendis M, Gainer V, Chueh HC, Churchill S, Kohane I. Serving the enterprise and beyond with informatics for integrating biology and the bedside (i2b2). *J Am Med Inform Assoc*. 2010;17(2):124–30.
  17. Simonneau G, Gatzoulis MA, Adatia I, Celermajer D, Denton C, Ghofrani A, Gomez Sanchez MA, Krishna Kumar R, Landzberg M, Machado RF, Olschewski H, Robbins IM, Souza R. Updated clinical classification of pulmonary hypertension. *J Am Coll Cardiol*. 2013;62(25 Suppl):D34–41.
  18. Ghofrani HA, D'armini AM, Grimminger F, Hoepfer MM, Jansa P, Kim NH, Mayer E, Simonneau G, Wilkins MR, Fritsch A, Neuser D, Weimann G, Wang C, CHEST-1 Study Group. Riociguat for the treatment of chronic thromboembolic pulmonary hypertension. *N Engl J Med*. 2013;369(4):319–29.
  19. Ghofrani HA, Simonneau G, D'armini AM, Fedullo P, Howard LS, Jaïs X, Jenkins DP, Jing ZC, Madani MM, Martin N, Mayer E, Papadakis K, Richard D, Kim NH, MERIT study investigators. Macitentan for the treatment of inoperable chronic thromboembolic pulmonary hypertension (MERIT-1): results from the multicentre, phase 2, randomised, double-blind, placebo-controlled study. *Lancet Respir Med*. 2017;5(10):785–94.
  20. Stubbs H, Lua S, Brewis M, Johnson M, Church C. COMPERA 2.0 risk stratification in medically-managed chronic thromboembolic pulmonary hypertension. *Eur Respir J*. 2022;60:22003113.
  21. Sahay S, Bhatt J, Beshay S, Guha A, Nguyen DT, Graviss EA, Nagueh SF. E-REVEAL Lite 2.0 scoring for early prediction of disease progression in pulmonary arterial hypertension. *Pulm Circ*. 2022;12(1):e12026.
  22. Boucly A, Weatherald J, Savale L, de Groote P, Cottin V, Prévot G, Chaouat A, Picard F, Horeau-Langlard D, Bourdin A, Jutant EM, Beurnier A, Jevnikar M, Jaïs X, Simonneau G, Montani D, Sitbon O, Humbert M. External validation of a refined 4-strata risk assessment score from the French pulmonary hypertension registry. *Eur Respir J*. 2022;59(6):2102419.

**How to cite this article:** Majeed RW, Wilkins MR, Howard L, Hassoun PM, Anthi A, Cajigas HR, Cannon J, Chan SY, Damonte V, Elwing J, Förster K, Frantz R, Ghio S, Al Ghouleh I, Hilgendorff A, Jose A, Juaneda E, Kiely DG, Lawrie A, Orfanos SE, Pepe A, Pepke-Zaba J, Sirenko Y, Swett AJ, Torbas O, Zamanian RT, Marquardt K, Michel-Backofen A, Antoine T, Wilhelm J, Barwick S, Krieb P, Fuenderich M, Fischer P, Gall H, Ghofrani H-A, Grimminger F, Tello K, Richter MJ, Seeger W. Pulmonary Vascular Research Institute GoDeep: a meta-registry merging deep phenotyping data from international PH reference centers. *Pulmonary Circulation*. 2022;12:e12123. <https://doi.org/10.1002/pul2.12123>