Learning a Comorbidity-Driven Taxonomy of Pediatric Pulmonary Hypertension

HHS Public Access

Author manuscript

Circ Res. Author manuscript; available in PMC 2018 August 04.

Published in final edited form as:

Circ Res. 2017 August 04; 121(4): 341–353. doi:10.1161/CIRCRESAHA.117.310804.

Learning a Comorbidity-Driven Taxonomy of Pediatric Pulmonary Hypertension

Mei-Sing Ong[1], **Mary P. Mullen**[2], **Eric D. Austin**[3], **Peter Szolovits**[4], **Marc Natter**[1,5], **Alon Geva**[1,6], **Tianxi Cai**[7], **Sek Won Kong**[1], and **Kenneth D. Mandl**[1]

[1]Computational Health Informatics Program, Harvard Medical School, Boston Children's Hospital, Boston MA

[2]Department of Cardiology, Harvard Medical School, Boston Children's Hospital, Boston MA

[3]Department of Pediatrics, Vanderbilt University Medical Center, Nashville, TN

[4]Computer Science and Artificial Intelligence Laboratory, Massachusetts Institute of Technology, Cambridge MA

[5]Department of Pediatrics, Massachusetts General Hospital, Boston MA

[6]Divison of Critical Care Medicine, Department of Anesthesiology, Perioperative, and Pain Medicine, Boston Children's Hospital, and Department of Anesthesia, Harvard Medical School, Boston MA

[7]Department of Biostatistics, Harvard School of Public Health, Boston MA

Abstract

Rationale—Pediatric pulmonary hypertension (PH) is a heterogeneous condition with varying natural history and therapeutic response. Precise classification of PH subtypes is therefore crucial for individualizing care. However, gaps remain in our understanding of the spectrum of PH in children.

Objective—We seek to study the manifestations of PH in children, and to assess the feasibility of applying a network-based approach to discern disease subtypes from comorbidity data recorded in longitudinal datasets.

Methods and Results—A retrospective cohort study comprising 6,943,263 children (<18 years of age) enrolled in a commercial health insurance plan in the US, between January 2010 and May 2013. A total of 1,583 (0.02%) children met the criteria for PH. We identified comorbidities significantly associated with PH compared with the general population of children without PH. A Bayesian comorbidity network was constructed to model the interdependencies of these comorbidities, and network clustering analysis was applied to derive disease subtypes comprising

Address correspondence to: Dr. Mei-Sing Ong, Boston Children's Hospital, 300 Longwood Ave, Boston MA 02115, Tel: 617-513-2755, Fax: 617-730-0267, Mei-Sing_Ong@hms.harvard.edu.

In May 2017, the average time from submission to first decision for all original research papers submitted to *Circulation Research* was 12.28 days.

DISCLOSURES

None.

subgraphs of highly connected comorbid conditions. A total of 186 comorbidities were found to be significantly associated with PH. Network analysis of comorbidity patterns captured most of the major PH subtypes with known etiological basis defined by the World Health Organization and Panama classifications. The analysis further identified a number of subtypes documented in only a few case studies, including rare subtypes associated with several well-described genetic syndromes.

Conclusions—Application of network science to model comorbidity patterns recorded in longitudinal datasets can facilitate the discovery of disease subtypes. Our analysis relearned established subtypes, thus validating the approach, and identified rare subtypes that are difficult to discern through clinical observations, providing impetus for deeper investigation of the disease subtypes that will enrich current disease classifications.

Keywords

Pulmonary hypertension; network science; comorbidity; pediatrics; clustering

Subject Terms

Pulmonary Hypertension; Pediatrics

INTRODUCTION

Classification of diseases has traditionally been shaped by expert consensus. As such, disease taxonomies are subject to the limits of existing knowledge and the biases of experts, and may not reflect the underlying disease pathophysiology. In a landmark report “Towards Precision Medicine”, the Institute of Medicine called for improved approaches to developing disease taxonomies that classify complex diseases based on the molecular and clinical data of individual patients [1]. Here, we developed a comorbidity-driven taxonomy of pediatric pulmonary hypertension (PH) using data derived from administrative claims.

It is well-established that PH is a heterogeneous condition, often with a progressive and potentially fatal course [2]. Effective management of PH relies on early and accurate diagnosis. However, diagnosing PH is challenging, as initial symptoms can be subtle and confounded by other pre-existing comorbidities. Patients with PH are often diagnosed late in the course when the pathologic changes are advanced and irreversible [3–7]. For example, one in five patients with pulmonary arterial hypertension (PAH), a subtype of PH, are diagnosed more than two years after symptom onset [4]. Given that the median survival for PAH is 2.8 years without treatment [8], delay in diagnosis likely worsens prognosis.

Accurate diagnosis of PH subtype is also critical, as available treatments and responses to treatments may differ among PH subtypes. Deaño et al reported a misdiagnosis rate of 33% among patients who were referred for care to tertiary level PH centers, and half of those who had started receiving medication treatment for PAH were found to either have a different PH subtype, or no PH at all [9]. Indeed, misdiagnosis and misuse of therapies exposes patients to the potential side effects of therapies without the supposed benefits; for instance, therapies which are efficacious for PAH (e.g., pulmonary vasodilator medications) may actually

worsen pulmonary hemodynamics in PH associated with left-heart disease, the most common cause of PH [10–11].

The goal to optimize the treatment of PH based on its pathophysiology has led to the development of a formal taxonomy of PH by the World Health Organization (WHO). The taxonomy (WHO Classification of PH) defines five distinct PH subtypes: PAH (Group 1 PH), PH due to left-heart disease (Group 2 PH), PH due to chronic lung disease and hypoxia (Group 3 PH), chronic thromboembolic PH (CTEPH) (Group 4 PH), and PH due to other multifactorial mechanisms (Group 5 PH) [2]. More recently, the realization that childhood-onset PH may have unique etiologies and association not often observed in adults further prompted the development of a new taxonomy for childhood-onset PH [12–14]. The taxonomy, known as the Panama classification (2011), highlighted a number of features more prominent in pediatric PH, including fetal and developmental origins of vascular disease [13].

While these efforts provide a framework for the diagnosis and treatment of PH, gaps remain in our understanding of the disease, especially among children. Due to the rarity of PH in children, comprehensive analysis of its clinical manifestations is challenging. To date, published data on pediatric PH have been limited to several registry-based and small cohort studies [3,15–18]. While these studies have greatly advanced our understanding of the disease, they may be subject to referral bias, and may not represent the full spectrum of pediatric PH cases [19–20]. Furthermore, since knowledge generated from these studies formed the bedrock of the expert consensus classifications, current taxonomies may not capture the full spectrum of the diverse manifestations of PH.

We therefore seek to enrich and extend the current classifications of PH by applying network science methodologies to the largest dataset of pediatric PH to date. Our goal is to facilitate improved recognition of clinically relevant patterns of disease manifestation, which can result in meaningful improvements in the timely and accurate diagnosis of PH.

The application of network science has transformed the study of biological systems [21]. Using network theory to model biological systems, nodes in the network represent biological entities (e.g. gene, protein, disease), and links between nodes represent the relationships between entities (e.g. transcriptional regulation, correlation in gene expression levels). The emergent properties of a network can provide insights into biological processes that cannot be elucidated by studying individual entities in isolation. For example, gene co-expression network studies have revealed novel genes involved in the pathogenesis of various diseases, while protein interaction networks have led to the delineation of new disease pathways. In addition to genomic and proteomic studies, comorbidity networks modeling disease co-occurrence have been shown to capture phenotypic differences between patients with different demographic backgrounds, disease progression and mortality [22]. Here, we apply for the first time a Bayesian network approach to characterize the comorbidity patterns in pediatric PH.

METHODS

This is a retrospective cohort study of an insured population. As shown in Figure 1, we (a) defined the comorbidity profiles of children with PH; (b) constructed a Bayesian comorbidity network; and (c) applied network clustering analysis to define disease subtypes.

Study population and data extraction

To form our study cohort, we examined claims from Aetna Inc., a major, nation-wide employer-provided health insurance plan in the United States between January 2010 and May 2013. Available information included dates of enrolment in the insurance program, outpatient visit claims and inpatient visit claims. Demographic data included sex and age. All encounters were coded with International Classification of Diseases, ninth revision (ICD-9) codes. The Boston Children's Hospital Institutional Review Board approved the study and granted a waiver of consent.

Subjects were drawn from 6,943,263 children (<18 years of age) enrolled in the insurer plan. A total of 1,583 children met the criteria for PH, defined as having two or more healthcare visits associated with PH (ICD-9 416.0, 416.8, 416.9) during the study period (Figure 2).

We systematically identified all comorbidities significantly associated with PH, compared with the general population of children without PH. We defined the presence of a comorbidity as having two or more care encounters related to the condition, identified using ICD-9 codes. This approach has been previously validated for a wide range of diseases [23–25]. For example, Quan et al. showed that patients with hypertension can be reliably identified with high accuracy (specificity: 95%, sensitivity: 73%) when the diagnoses were defined as having two or more claims for hypertension within three years [24]. We further conducted an additional sensitivity analysis, applying a more stringent algorithm that considered only diagnoses with four or more encounter visits. Using chi-square statistic, we calculated the odds ratio with corresponding 95% confidence intervals (95% CI) and two tailed p-values to measure the strength of the association between each comorbidity and PH. The α level of 0.05 was used to declare statistical significance. Bonferroni correction was applied to control for 767 comparisons with the adjusted α level being 0.000065.

Bayesian comorbidity network

We developed a Bayesian network to model the inter-dependencies of comorbidities in children with PH. We included in the model all comorbidities found to be significantly associated with PH compared with the general population of children without PH (Supplementary Table B1a). We excluded comorbidities that occurred in fewer than four PH patients, since a small number of observations would not suffice to distinguish between true and spurious correlations. Schematically, the model is represented as a directed graph, with nodes representing comorbidities, and edges denoting statistically dependent relations among comorbidities [26]. An edge from node C_i to C_j can be interpreted as the presence of C_i "influences" the occurrence of C_j. The absence of a link between two nodes signifies conditional independence. Detailed description of the methods is provided in Online Supplement Section A1.

Network clustering analysis

To define PH subtypes, we partitioned the network into subgraphs comprising highly connected comorbidities using a strict partitioning rule whereby each comorbidity belongs to exactly one cluster. The graph partitioning process involves merging nodes agglomeratively, using a random walk clustering approach [27]. Detailed description of the methods is provided in Online Supplement Section A2.

Evaluation

Review by experts evaluated our approach by checking for the identification of established PH subtypes. Accordingly, each comorbidity cluster was assigned a WHO and Panama classification subtype that best describes the cluster. For example, a comorbidity cluster comprising portal hypertension and the associated conditions would be classified as WHO Group 1 (PAH associated with portal hypertension), and Panama Group 10 (pediatric pulmonary vascular disease associated with other system disorders). Classification was first performed by one researcher (MSO), and then validated by two pediatric PH experts (MH and EA); inter-rater agreement was quantified using Cohen's kappa statistic and discrepancies were resolved through consensus. The WHO and Panama classifications comprise five and ten subtypes, respectively; the WHO classification further categorizes some of the five major subtypes into 26 "minor" subtypes. (Figure 1C). To ascertain the sensitivity of our analysis, we measured the percentage of subtypes defined in WHO/Panama classifications that were captured by the network. We further conducted a literature review to evaluate network-derived clusters not described in either WHO or Panama classifications to assess if published evidence supported the co-occurrence of PH with conditions captured by each cluster.

Secondary analysis

To form the study cohort, we considered children who had two or more healthcare visits for PH. However, given that PH can often be miscoded [28] or misdiagnosed [9, 29], we may have inadvertently included patients who did not have PH. Definitive diagnosis of PH requires an elevated mean pulmonary arterial pressure of 25 mmHg at rest measured by right heart catheterization (RHC). Hence, to address potential coding and diagnostic ambiguities, we performed a secondary analysis that considered only children who underwent RHC, in addition to having two or more healthcare visits for PH during the study period. A separate Bayesian comorbidity network was developed for this patient subgroup, and the differences in comorbidities and network-derived clusters were assessed.

RESULTS

A total of 6,943,263 children were enrolled in the insurance plan and 1,583 (0.02%) had two or more diagnosis codes for PH. The 6,940,062 children without any diagnosis codes for PH formed the control population for this study. A total of 186 comorbidities satisfied our study inclusion criteria (Supplementary Table B1a). When a more stringent algorithm for identifying diagnoses was applied, defined as having four or more healthcare visits during the study period, the association between PH and these comorbidities remained statistically significant, with the exception of pyogenic granuloma (Supplementary Table B1b).

The comorbidity burden of children with PH was substantial (Figure 3). The mean and median number of comorbid conditions in individuals with PH were 7 and 8, respectively.

Network analysis

Figure 4 depicts the Bayesian comorbidity network learned from the dataset. The inferred network comprised 365 relations, described in further details in Supplementary Table B2a.

Detection of well-established subtypes

Cluster analysis of the comorbidity network identified all five major subtypes (sensitivity: 100%; kappa score: 100%) and 19 of 26 minor subtypes (sensitivity: 73%; kappa score: 96%) defined in the WHO classification (Table 1), and nine of ten subtypes defined in the Panama classification (sensitivity: 90%; kappa score: 90%) (Table 2), with a few anticipated exceptions. For example, in the absence of pedigree and genetic data, we were unable to discern the various forms of heritable PAH, with the exception of PH in association with hereditary hemorrhagic telangiectasia (HHT), a condition known to associate with PAH and linked to pathogenic variants in ALK1 and ENG genes [30]. The identification of pathogenic drugs and toxins associated with PAH is beyond the scope of this study. We were also unable to detect PH caused by chronic exposure to high altitude: the diagnostic code for the condition is non-specific (ICD-9 993.2: “other and unspecified effects of high altitude”), and was not assigned to any of the patients in the study dataset. The imprecision of ICD codes also precluded differentiation between left ventricular systolic and diastolic dysfunctions. Finally, we were unable to discern subtypes associated with HIV infection or schistosomiasis, since none of the PH patients in our US-based, privately-insured claims dataset had billing codes for these conditions.

Detection of rare subtypes

Our analysis detected known rare associations with PH (Table 3). An example is the co-occurrence of PH with juvenile idiopathic arthritis (JIA) and hemophagocytic syndrome. The clustering occurrence is not surprising given that PAH has been reported in several patients with systemic-onset JIA, particularly in association with macrophage activation syndrome (MAS) [31–32]. It has been hypothesized that PAH may be caused by exposures to IL-1 and IL-6 inhibitors used for treating systemic-onset JIA and MAS [33]; however, the underlying biology of this association remains unknown. In our dataset, of 18 patients with both JIA and hemophagocytic syndrome, four patients developed PH, signifying the potential importance of this comorbidity pattern.

The cluster comprising glycogen storage disorder (GSD), hereditary muscular dystrophy and cardiomyopathy typifies type 2 GSD. While type 1 GSD has been linked to PH, the relationship between PH and type 2 GSD is less studied. A case report noted the development of PH resulting from respiratory muscular atrophy and alveolar hypoventilation caused by type 2 GSD [34]. Another report documented severe pulmonary veno-occlusive disease (PVOD) in a patient with type 2 GSD [35].

Another cluster captures the characteristic features of heterotaxy syndrome, including situs inversus and congenital spleen anomaly. Of the 31 patients in our dataset with these

conditions, four developed PH. Several limited case reports documented the disease pattern in the setting of cardiac defects and pulmonary complications [36–37].

We further identified several rare genetic disorders among the PH population, including Cri-du-chat, Turner and Prader-Willi syndromes. Case reports have documented the co-occurrence of PH in children with Cri-du-chat [38–39] and Turner syndrome [40–41]. PH in these patients may be caused by underlying congenital heart disease. In patients with Prader-Willi syndrome, obstructive sleep apnea and other obesity-related comorbidites may have contributed to the development of PH. In our dataset, six (1.8%) of 329 patients with Prader-Willi syndrome developed PH. However, literature review yielded only one case report documenting a sudden death secondary to PH in a child with Prader-Willi syndrome [42], indicating that the risk of PH in these patients may be under-recognized.

While some clusters clearly describe syndromes with highly specific and/or rare comorbidities, other clusters contain unusual combinations of relatively more common conditions, which may represent unrecognized syndromes and generate new hypotheses. For example, the cluster comprising adrenogenital disorder, microcephaly, and adrenal hypofunction may represent Smith-Lemli-Opitz syndrome (SLOS), a rare condition caused by deficiency of 7-dehydrocholesterol reductase. Two potential etiologies would support the association between SLOS and PH. First, persistent pulmonary hypertension in newborn (PPHN) in SLOS has been documented in a patient with altered expression of caveolin-1 [43], suggesting that caveolae-dependent signaling may be responsible for the pathogenesis of PH. This hypothesis was further strengthened in a recent study demonstrating an association between mutations in CAV-1 and PAH through whole exome sequencing [44]. Second, cardiorespiratory problems can occur in individuals with SLOS, secondary to malformations of the heart or respiratory tract [45]; these conditions may contribute to the development of PH in patients with SLOS.

We further identified a cluster containing comorbidities suggestive of Ehlers Danlos syndrome (EDS), a group of genetically-determined connective-tissue disorders (CTD) presenting with musculoskeletal dysfunction, encompassing a spectrum of symptoms such as hyperextensible skin, spontaneous ecchymoses, fragile vessels, and hypermobile joints with oftentimes secondary gait abnormalities. Spontaneous vascular or visceral rupture is a feature typifying vascular EDS, a rare subtype of EDS caused by mutations in the COL3A1 gene that result in increased fragility of connective tissue. In our study population, two of 11 children who had co-diagnoses of ‘spontaneous ecchymoses’, ‘disorder of muscle or ligament’, and ‘gait abnormality’ were diagnosed with PH, raising the hypothesis for an association between vascular EDS and PH. The co-occurrence of EDS and mild PH has been documented in two patients [46]; abnormalities in the pulmonary vasculature might also have predisposed these patients to PH.

The association of PH with intestinal malabsorption and perinatal digestive system disorders may capture a rare treatable form of PH that is linked to chronic nutritional deficiencies. Severe vitamin deficiencies have been documented in patients with PH, whereby repletion of these vitamins led to the resolution of PH. A recent study showed a high prevalence of vitamin D deficiency in patients with PH, and a significant inverse correlation between

vitamin D serum levels and disease severity [47]. Furthermore, thiamine deficiency resulting in cardiovascular beriberi has also been reported as a rare reversible cause of PH [48]. PH associated with vitamin C deficiency in the setting of pulmonary complications has been documented in several patients [49–50]. PH in these patients may be a result of reduced synthesis of nitric oxide, a potent mediator of vascular muscle relaxation. Vitamin C is also essential for the regulation of hypoxia-inducible family of transcription factors; deficiency in vitamin C can result in the inactivation of HIF activity, leading to deleterious pulmonary vasoconstriction and PH.

Discovery of unknown subtypes

Several network-derived comorbidity clusters do not fall into any of the categories in the WHO and Panama classifications. Of note, a number of these clusters are linked to neurological defects not commonly thought to be associated with PH, including encephalocele, hydrocephalus, microcephaly, periventricular leukomalacia, and congenital brain reduction deformities (Table 4). It is well-established that children with severe neurological impairments are predisposed to respiratory problems that occur as a direct consequence of the underlying disability. For example, oropharyngeal motor problems associated with neurological dysfunctions can lead to recurrent aspiration and pneumonia [51]. Chiari malformation associated with hydrocephalus can cause both maldevelopment of the brain stem respiratory control centers and central sleep apnea [52]. Neurological impairments are also common among children requiring mechanical ventilation at the intensive care unit; PH in these patients may be secondary to mechanical ventilation management. While we are unable to ascertain the causes of PH observed in our study cohort, our analysis suggests that the association of neurological defects with PH may be under-recognized and deserves further characterization.

Secondary analysis: RHC cohort

Of the 1,583 patients included in the primary study cohort of children, 308 underwent RHC during the study period. Subject attrition was likely a result of the limited observation period captured in our dataset, spanning a period of only three years; thus, patients who underwent RHC outside the study observation period would have been missed. Furthermore, published studies have consistently reported that many PH patients do not receive RHC as part of their diagnostic workup, despite being the diagnostic gold standard for PH [9, 29, 53].

Of the 186 comorbidities found to be significantly associated with PH in the primary study cohort, 168 were also significantly associated with PH in the RHC cohort, of which 119 were observed in four or more patients with PH (Supplementary Table B1c). A larger proportion of patients in the RHC cohort had congenital heart disease (CHD), compared with the primary cohort (86.0% [n=1361] vs 67.9% [n=209]; $p<0.0001$). This may reflect a higher utilization of RHC among PH patients treated by cardiologists [29]. There was also a substantial drop in the proportion of patients who had early-life respiratory conditions known to be associated with PH, including PPHN (25.6% [n=406] vs 11.7% [n=36]; $p<0.0001$), congenital cystic lung (11.4% [n=180] vs n=0; $p<0.0001$), and diaphragmatic hernia (3.9% [n=61] vs 0.6% [n=2]; $p=0.0041$).

A separate Bayesian comorbidity network was developed for the RHC cohort. The input of the network comprised 119 comorbidities that were significantly associated with PH, and were observed in four or more patients in the RHC cohort. The inferred network contained 200 relations, described in further details in Supplementary Table B2b.

Cluster analysis of the comorbidity network identified a subset of the subtypes derived from the primary study cohort. These included all the major subtypes (sensitivity: 100%; kappa score: 100%) and 17 of 26 minor subtypes (sensitivity: 65%; kappa score: 100%) defined in the WHO classification (Supplementary Table B3a), and eight of ten subtypes defined in the Panama classification (sensitivity: 80%; kappa score: 100%) (Supplementary Table B3b). Subtypes that were not captured included PH associated with HHT, CTD, systemic vasculitis, GSD – these conditions were significantly associated with PH in the RHC cohort, but were observed in less than four patients, and were therefore excluded from the network analysis. Rare genetic disorders discerned in the primary analysis, such as Cri-du-chat and Prader-Willi syndromes, were also not captured in the network analysis due to the small number of patients with these conditions.

DISCUSSION

We demonstrated that comorbidity patterns of patients with PH captured in a Bayesian network can be stratified into subtypes that are biologically and clinically informative. Our algorithmic methods automatically relearned most of the major PH subtypes with known etiological basis defined by the WHO classification. While the similarity of the derived network structure to current taxonomy of PH provides face validity to the approach, it also offers some novel insights.

Specifically, the network approach enriches the current classification of PH. First, it captured several subtypes documented in only a few case studies for which evidence for systematic association remains lacking. This both validates the approach and provides impetus for deeper investigation of the disease subtypes. Clinicians should consider PH when dealing with patients with relevant conditions, while studies exploring the molecular and biologic connections may reveal important insights. Our analysis also identified rare subtypes with findings consistent with several well-described genetic syndromes. While we were unable to validate if these patients indeed had the pathogenic mutations in the absence of genetic data, disease patterns identified through network analysis suggest future research in these areas may accelerate subtype discovery and relevance to PH more broadly. In the same way in which novel genetic associations in PH stimulate new avenues of research, so too may novel phenotypic associations prompt important discoveries related to disease susceptibility, and perhaps resiliency.

To construct the comorbidity network, we applied Bayesian model averaging technique to find a network model that best fits the underlying data. The approach is uniquely suited to accommodate the inherent uncertainties of biological processes, and to minimize the effects of noise in the data. In maximizing specificity, however, other subtypes may have been missed. Furthermore, in defining comorbidity clusters, we have applied a strict partitioning rule, whereby each comorbidity belongs exactly to one cluster. While this approach produces

a model that is easier to interpret, the full expression of subtypes may not have been captured. As shown in Figure 4, many comorbidities are linked to comorbidities belonging to another cluster. Of note, a majority of nodes are connected either directly or indirectly to congenital heart disease, the most common comorbidity in our study cohort. Shared features among multiple clusters may also reflect the overlapping phenotypes of PH, an increasingly recognized phenomenon [14]. Future research should explore methods that would facilitate the delineation of subtypes with overlapping manifestations and etiologies.

There are several limitations to our study. First, because our study relied on administrative claims, the diagnoses coded for billing purposes may not reflect actual comorbidities in the patients. To improve case identification specificity, we included only diagnoses with two or more encounter visits – an algorithm that has been validated in previous studies [23–25]. We further performed a sensitivity analysis with a more stringent case identification algorithm, considering diagnoses with four or more encounter visits during the study period; the association between PH and the identified comorbidities remained significant. While it may not be possible to fully address diagnostic coding inaccuracies in administrative claims, we were able to discern comorbidity relations and derive subtypes that are biologically meaningful, thus lending support to the validity of the study approach. A further strategy we used to reduce uncertainties in our analysis is the exclusion of comorbidities that occurred in fewer than four patients, and network relations with low probabilistic strengths (Supplementary Methods A1). In doing so, we may not have captured all the comorbidities and subtypes that were present in our study cohort.

A further limitation of our study lies in the difficulty in ascertaining the study cohort – in the absence of RHC results, we were unable to definitively confirm a diagnosis of PH. Nonetheless, the number of children with PH identified in the primary analysis was within the estimated population prevalence of PH [17]. We have also conducted a secondary analysis of children who underwent RHC and were diagnosed with PH. However, given the limited study period, and the persistent underuse of recommended diagnostic tests in contemporary clinical practice [53–54], restricting the study population results in the exclusion of several important subtypes. Nonetheless, the ability to reproduce similar subtypes based on the RHC cohort further validates the approach.

Claims data also have limited resolution, and since comorbidities associated with different PH subtypes can often coexist, we were unable to ascertain the PH subtypes of individual patients based on concomitant conditions alone. For example, patients with systemic sclerosis can have both PAH and pulmonary fibrosis; differentiating PAH from PH associated with pulmonary fibrosis can only be achieved through RHC. Finally, while our claims-based dataset formed the largest dataset on pediatric PH to date, we were unable to capture all PH subtypes seen globally, including those associated with HIV and schistosomiasis.

While there are limitations to claims data, the availability of a large number of patients makes it possible to study the relationships between rare diseases, which may not have been observable in traditional studies involving chart reviews or surveys. Further, claims data are systematically collected, and provide longitudinal information that crosses facilities,

geographical locations and population demographics, thereby enhancing the generalizability of the research and limiting selection biases.

CONCLUSIONS

Data-driven discovery of disease phenotypes has enormous potential for advancing our understanding of complex diseases. By modeling the complex interactions of symptoms governing disease subtypes recorded in longitudinal datasets, rare subtypes that are difficult to discern through clinical observations can be identified. Further advances linking disease subtypes to therapeutic response, disease outcomes, as well as the molecular profiles of individual subtypes, will provide impetus for the development of more effective and targeted therapies.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

We thank the PPHNet and NHLBI Pediatric Pulmonary Vascular Disease Outcomes Bioinformatics Clinical Coordinating Center Investigators for their valuable comments. Michelle Heinz (Boston Children's Hospital) is gratefully acknowledged for her assistance in coordinating the study.

SOURCES OF FUNDING

Research reported in this publication was supported by Aetna Inc., and the National Heart, Lung, And Blood Institute of the National Institutes of Health under Award Number U01HL121518. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health. MSO and AG are supported by the National Institutes of Health (T15LM007092, T32HD040128 and L40HL133929). The funding bodies played no role in the design and conduct of the study; collection, management, analysis, and interpretation of the data; and preparation, review, or approval of the manuscript; and decision to submit the manuscript for publication.

Nonstandard Abbreviations and Acronyms

CTD	Connective tissue disease
CTEPH	Chronic thromboembolic pulmonary hypertension
GSD	Glycogen storage disorder
HHT	Hereditary hemorrhagic telangiectasia
JIA	Juvenile idiopathic arthritis
MAS	Macrophage activation syndrome
PAH	Pulmonary arterial hypertension
PPHN	Pulmonary hypertension in newborn
PH	Pulmonary hypertension
PVOD	Pulmonary veno-occlusive disease

RHC	Right heart catheterization
SLOS	Smith-Lemli-Opitz syndrome
WHO	World Health Organization

References

1. National Research Council Committee on A Framework for Developing a New Taxonomy of Disease. Toward precision medicine: building a knowledge network for biomedical research and a new taxonomy of disease. National Academies Press; 2011.
2. 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:D34–41. [PubMed: 24355639]
3. McGoon MD, Benza RL, Escribano-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:D51–9. [PubMed: 24355642]
4. George MG, MG, Schieb LJ, Ayala C, Talwalkar A, Levant S. Pulmonary hypertension surveillance: United States, 2001 to 2010. Chest. 2014; 146:476–95. [PubMed: 24700091]
5. Brown LM, Chen H, Halpern S, Taichman D, McGoon MD, Farber HW, Frost AE, Liou TG, Turner M, Feldkircher K, Miller DP, Elliott CG. Delay in recognition of pulmonary arterial hypertension: factors identified from the REVEAL Registry. Chest. 2011; 140:19–26. [PubMed: 21393391]
6. Strange G, Gabbay E, Kermeen F, Williams T, Carrington M, Stewart S, Keogh A. Time from symptoms to definitive diagnosis of idiopathic pulmonary arterial hypertension: the delay study. Pulm Circ. 2013; 31:89–94.
7. Wilkens H, Grimminger F, Hoeper M, Stahler G, Ehlken B, Plesnila-Frank C, et al. Burden of pulmonary arterial hypertension in Germany. Respir Med. 2010; 104:902–10. [PubMed: 20149617]
8. D'Alonzo GE, Barst RJ, Ayres SM, Bergofsky EH, Brundage BH, Detre KM, Fishman AP, Goldring RM, Groves BM, Kernis JT, et al. Survival in patients with primary pulmonary hypertension. Results from a national prospective registry. Ann Intern Med. 1991; 115:343–9. [PubMed: 1863023]
9. Deaño RC, Glassner-Kolmin C, Rubenfire M, Frost A, Visovatti S, McLaughlin VV, Gomberg-Maitland M. Referral of patients with pulmonary hypertension diagnoses to tertiary pulmonary hypertension centers: the multicenter RePHerral study. JAMA Intern Med. 2013; 173:887–93. [PubMed: 23568223]
10. Packer M, McMurray J, Massie BM, Caspi A, Charlon V, Cohen-Solal A, Kiowski W, Kostuk W, Krum H, Levine B, Rizzon P, Soler J, Swedberg K, Anderson S, Demets DL. Clinical effects of endothelin receptor antagonism with bosentan in patients with severe chronic heart failure: results of a pilot study. J Card Fail. 2005; 11:12–20. [PubMed: 15704058]
11. Haywood GA, Sneddon JF, Bashir Y, Jennison SH, Gray HH, McKenna WJ. Adenosine infusion for the reversal of pulmonary vasoconstriction in biventricular failure. A good test but a poor therapy. Circulation. 1992; 86:896–902. [PubMed: 1516202]
12. Barst RG, Ertel SI, Beghetti M, Ivy DD. Pulmonary arterial hypertension: a comparison between children and adults. Eur Respir J. 2011; 37:665–77. [PubMed: 21357924]
13. Ivy DD, Abman SH, Barst RJ, Berger RM, Bonnet D, Fleming TR, Haworth SG, Raj JU, Rosenzweig EB, Schulze Neick I, Steinhorn RH, Beghetti M. Pediatric pulmonary hypertension. J Am Coll Cardiol. 2013; 62:D117–26. [PubMed: 24355636]
14. Cerro MR, Abman S, Diaz G, Freudenthal AH, Freudenthal F, Harikrishnan S, Haworth SG, Ivy D, Lopes AA, Raj JU, Sandoval J, Stenmark K, Adatia I. A consensus approach to the classification of pediatric pulmonary hypertensive vascular disease: Report from the PVRI pediatric taskforce, Panama 2011. Pulm Circ. 2011; 1:286–298. [PubMed: 21874158]
15. Berger RM, et al. Clinical features of paediatric pulmonary hypertension: a registry study. Lancet. 2012; 379:537–46. [PubMed: 22240409]

16. Badesch DB, Raskob GE, Elliott CG, Krichman AM, Farber HW, Frost AE, Barst RJ, Benza RL, Liou TG, Turner M, Giles S, Feldkircher K, Miller DP, McGoon MD. Pulmonary arterial hypertension: baseline characteristics from the REVEAL registry. Chest. 2010; 137:376–87. [PubMed: 19837821]
17. van Loon RL, Roofthooft MT, Hillege HL, ten Harkel AD, van Osch-Gevers M, Delhaas T, Kapusta L, Strengers JL, Rammeloo L, Clur SA, Mulder BJ, Berger RM. Pediatric pulmonary hypertension in the Netherlands: epidemiology and characterization during the period 1991 to 2005. Circulation. 2011; 124:1755–64. [PubMed: 21947294]
18. Moledina S, Hislop AA, Foster H, Schulze-Neick I, Haworth SG. Childhood idiopathic pulmonary arterial hypertension: a national cohort study. Heart. 2010; 96:1401–6. [PubMed: 20406768]
19. Fraisse A, Jais X, Schleich JM, di Filippo S, Maragnès P, Beghetti M, Gressin V, Voisin M, Dauphin C, Clerson P, Godart F, Bonnet D. Characteristics and prospective 2-year follow-up of children with pulmonary arterial hypertension in France. Arch Cardiovasc Dis. 2010; 103:66–74. [PubMed: 20226425]
20. Miller DP, Gomberg-Maitland M, Humbert M. Survivor bias and risk assessment. Eur Respir J. 2012; 40:530–2. [PubMed: 22941543]
21. Chan SY, Loscalzo JL. The emerging paradigm of network medicine in the study of human disease. Circ Res. 2012; 111:359–374. [PubMed: 22821909]
22. Hildago CA, Blumm N, Barabasi AL, Christakis NA. A dynamic network approach for the study of human phenotypes. PLoS Computational Biology. 2009; 5:e1000353. [PubMed: 19360091]
23. Lix, L., Yogendran, M., Burchill, C., Metge, C., McKeen, N., Moore, D., Bond, R. Defining and validating chronic diseases: an administrative data approach. Winnipeg: Manitoba Centre for Health Policy; 2006.
24. Quan H, Khan N, Hemmelgarn BR, Tu K, Chen G, Campbell N, Hill MD, Ghali WA, McAlister FA. Validation of a case definition to define hypertension using administrative data. Hypertension. 2009; 54:1423–8. [PubMed: 19858407]
25. Rector TS, Wickstrom SL, Shah M, Thomas Greeenlee N, Rheault P, Rogowski J, Freedman V, Adams J, Escarce JJ. Specificity and sensitivity of claims-based algorithms for identifying members of Medicare+Choice health plans that have chronic medical conditions. Health Serv Res. 2004; 39:1839–57. [PubMed: 15533190]
26. Pearl, J. Causality: models, reasoning and inference. Cambridge: MIT press; 2000.
27. Pons, P., Latapy, M. Computer and Information Sciences-ISCIS 2005. Springer; Berlin Heidelberg: 2005. Computing communities in large networks using random walks; p. 284-293.
28. Link J, Glazer C, Torres C, Chin K. International Classification of Diseases Coding Changes Lead to Profound Declines in Reported Idiopathic Pulmonary Arterial Hypertension Mortality and Hospitalizations: Implications for Database Studies. Chest. 2011; 139:497–504. [PubMed: 20724737]
29. Duarte AG, Lin YL, Sharma G. Incidence of right heart catheterization in patients initiated on pulmonary arterial hypertension therapies: A population-based study. J Heart Lung Transplant. 2017; 36:220–226. [PubMed: 27595194]
30. Harrison RE, Flanagan JA, Sankelo M, Abdalla SA, Rowell J, Machado RD, et al. Molecular and functional analysis identifies ALK-1 as the predominant cause of pulmonary hypertension related to hereditary haemorrhagic telangiectasia. J Med Genet. 2003; 40:865–71. [PubMed: 14684682]
31. Kimura Y, Weiss JE, Haroldson KL, Lee T, Punaro M, Oliveira S, et al. Pulmonary hypertension and other potentially fatal pulmonary complications in systemic juvenile idiopathic arthritis. Arthritis Care Res. 2013; 65:745–752.
32. Li EK, Tam LS. Pulmonary hypertension in systemic lupus erythematosus: Clinical association and survival in 18 patients. J Rheumatol. 1999; 26:1923–1929. [PubMed: 10493670]
33. Humbert M, Monti G, Brenot F, Sitbon O, Portier A, Grangeot-Keros L, et al. Increased interleukin-1 and interleukin-6 serum concentrations in severe primary pulmonary hypertension. Am J Respir Crit Care Med. 1995; 151:1628–31. [PubMed: 7735624]
34. Inoue S, Nakamura T, Hasegawa K, Tadaoka S, Samukawa M, Nezuo S, et al. Pulmonary hypertension due to glycogen storage disease type II (Pompe's disease): a case report. J Cardiol. 1989; 19:323–32. [PubMed: 2810049]

35. Kobayashi H, Shimada Y, Ikegami M, Kawai T, Sakurai K, Urashima T, et al. Prognostic factors for the late onset Pompe disease with enzyme replacement therapy: from our experience of 4 cases including an autopsy case. Mol Genet Metab. 2010; 100:14–9. [PubMed: 20202878]
36. Brandenburg VM, Krueger S, Haage P, Mertens P, Riehl J. Heterotaxy syndrome with severe pulmonary hypertension in an adult patient. South Med J. 2002; 95:536–8. [PubMed: 12005012]
37. Yousuf T, Kramer J, Jones B, Keshmiri H, Dia M. Pulmonary hypertension in a patient with congenital heart defects and heterotaxy syndrome. Ochsner J. 2016; 16:309–11. [PubMed: 27660582]
38. Hills C, Moller JH, Finkelstein M, Lohr J, Schimmenti L. Cri du Chat syndrome and congenital heart disease: A review of previously reported cases and presentation of an additional 21 cases from the Pediatric Cardiac Care Consortium. Pediatrics. 2006; 117:e924–7. [PubMed: 16585274]
39. Levy B, Dunn TM, Kern JH, Hirschhorn K, Kardon NB. Delineation of the dup5q phenotype by molecular cytogenetic analysis in a patient with dup5q/del 5p (cri du chat). Am J Med Genet. 2002; 108:192–7. [PubMed: 11891684]
40. Bechtold SM, Dalla Pozza R, Becker A, Meidert A, Döhlemann C, Schwarz HP. Partial anomalous pulmonary vein connection: an underestimated cardiovascular defect in Ullrich-Turner syndrome. Eur J Pediatr. 2004; 163:158–62. [PubMed: 14716560]
41. Tinker A, Schofield UJ. Severe pulmonary hypertension in Turner syndrome. Br Heart J. 1989; 62:74–77. [PubMed: 2757877]
42. Bakker B, Maneatis T, Lippe B. Sudden death in Prader-Willi Syndrome: brief review of five additional cases. Horm Res. 2007; 67:203–204. [PubMed: 17164585]
43. Katheria AC, Masliah E, Benirschke K, Jones KL, Kim JH. Idiopathic persistent pulmonary hypertension in an infant with Smith-Lemli-Opitz syndrome. Fetal Pediatr Pathol. 2010; 29:373–9. [PubMed: 21043560]
44. Austin ED, Ma L, LeDuc C, Berman Rosenzweig E, Borczuk A, Phillips JA 3rd, et al. Whole exome sequencing to identify a novel gene (caveolin-1) associated with human pulmonary arterial hypertension. Circ Cardiovasc Genet. 2012; 5:336–43. [PubMed: 22474227]
45. Nowaczyk, M. GeneReviews. 2013. Smith-Lemli-Opitz syndrome.
46. McQuillan BM, Picard MH, Leavitt M, Weyman AE. Clinical correlates and reference intervals for pulmonary artery systolic pressure among echocardiographically normal subjects. Circulation. 2001; 104:2797–2802. [PubMed: 11733397]
47. Tashtoush B, Mehta J, Benavides J, Memarpour R, Kachur S, Gleason JB, Hadeh A, Smolley L, Rahaghi F. The Prevalence of Vitamin D Deficiency in Pulmonary Hypertension Patients and Its Correlation with Parameters of Pulmonary Hypertension Severity. Am J Respir Crit Care Med. 2016; 193:A7358.
48. Goto Y, Awata M, Uematsu M, Yano M, Morozumi T, Kotani J, Watanabe T, Onishi T, Iida O, Sera F, Nanto S, Nagata S. Soshin beriberi complicating severe pulmonary hypertension: a case report. J Cardiol. 2007; 49:361–5. [PubMed: 17633575]
49. Duvall MG, Pikman Y, Kantor DB, Ariagno K, Summers L, Sectish TC, Mullen MP. Pulmonary hypertension associated with scurvy and vitamin deficiencies in an autistic child. Pediatrics. 2013; 132:e1699–703. [PubMed: 24190688]
50. Kupari M, Rapola J. Reversible pulmonary hypertension associated with vitamin C deficiency. Chest. 2012; 142:225–7. [PubMed: 22796843]
51. Seddon PC, Khan Y. Respiratory problems in children with neurological impairment. Arc Dis Child. 2003; 88:75–8.
52. Dauvilliers Y, Stal V, Abril B, Coubes P, Bobin S, Touchon J, Escourrou P, Parker F, Bourgin P. Chiari malformation and sleep related breathing disorders. J Neurol Neurosurg Psychiatry. 2007; 78:1344–8. [PubMed: 17400590]
53. McLaughlin VV, Langer A, Tan M, et al. Contemporary trends in the diagnosis and management of pulmonary arterial hypertension: an initiative to close the care gap. Chest. 2013; 143:324–32. [PubMed: 22910999]
54. Douwes JM, Hump T, Bonnet D, Beghetti M, Ivy DD, Berger RM. TOPP Investigators. Acute vasodilator response in pediatric pulmonary arterial hypertension: current clinical practice from the TOPP registry. JACC. 2016; 67:1312–1323. [PubMed: 26988953]

NOVELTY AND SIGNIFICANCE

What Is Known?

- Pulmonary hypertension (PH) is a complex disease with varying disease course and treatment response.
- Formal taxonomies of PH have been developed to guide diagnosis and targeted treatment based on the underlying disease pathophysiology.
- Gaps remain in our knowledge of the diverse forms of PH, especially in children.

What New Information Does This Article Contribute?

- Application of network science to model disease patterns captured in longitudinal datasets provides an unbiased approach to discovering disease subtypes.
- The approach identified rare and novel disease patterns in pediatric PH, providing impetus for deeper investigation that will enrich current disease classifications.

Classification of diseases has traditionally been shaped by expert consensus. As such, disease taxonomies are subject to the limits of existing knowledge and the biases of experts, and may not reflect the underlying disease pathophysiology. Here, we establish for the first time the feasibility of applying network science to drive data-driven discovery of disease subtypes in pediatric PH. By modeling the complex interactions of disease symptoms recorded in longitudinal datasets, the approach automatically relearned PH subtypes with known etiological basis, and further identified novel and rare disease patterns that are difficult to discern through clinical observations. Knowledge derived from the analysis will facilitate improved diagnosis of PH. Further advances linking disease subtypes to therapeutic response, disease outcomes, as well as the molecular profiles of individual subtypes, will provide impetus for the development of more effective and targeted therapies. While the current study focuses on pediatric PH, the proposed method has enormous potential for advancing our understanding of other complex diseases.

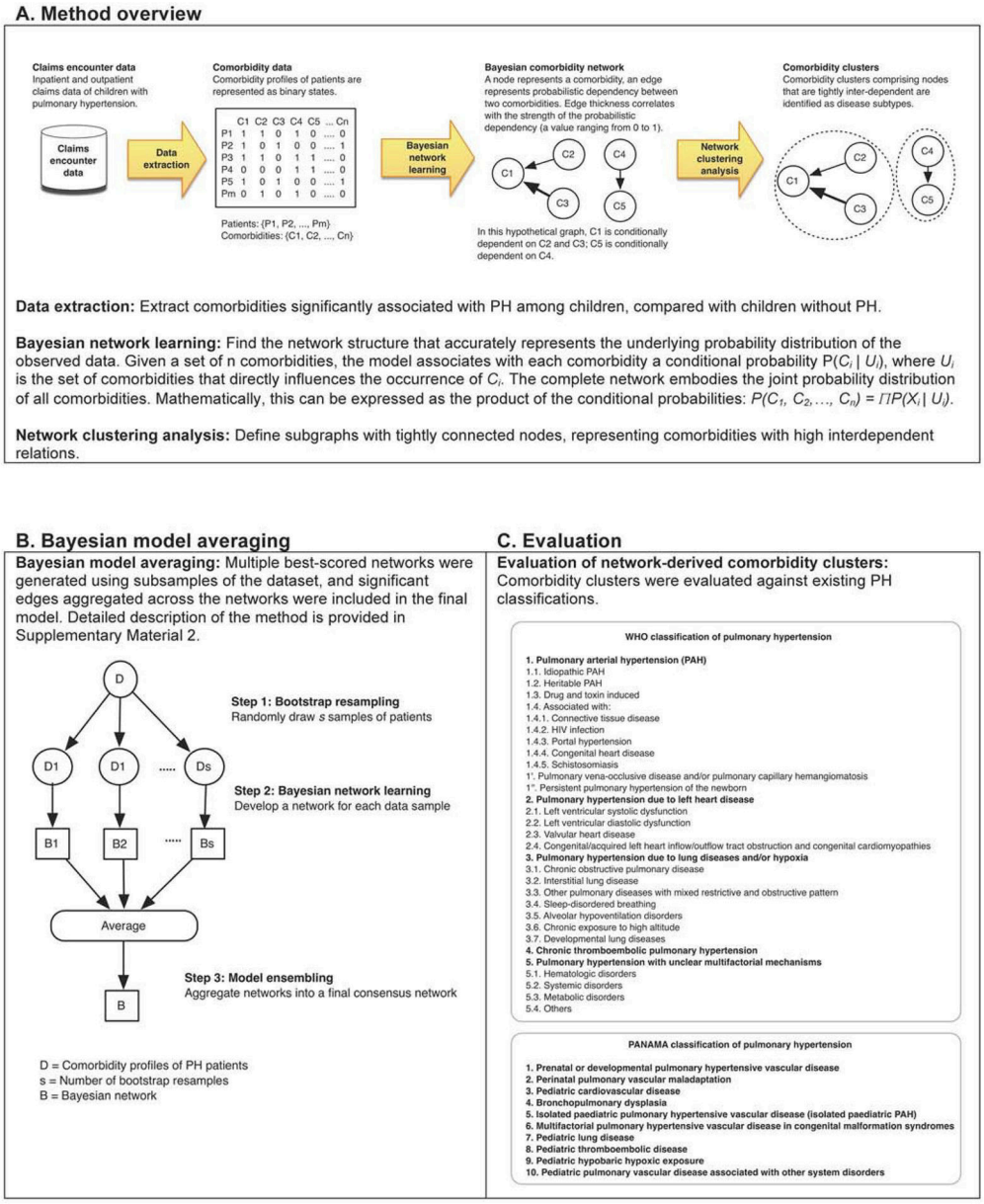


Figure 1.
Study methods

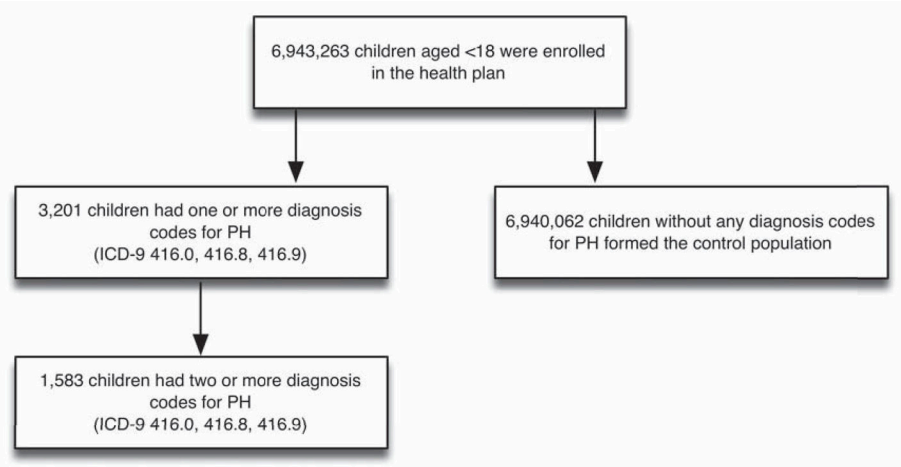


Figure 2.
Study participants selection criteria

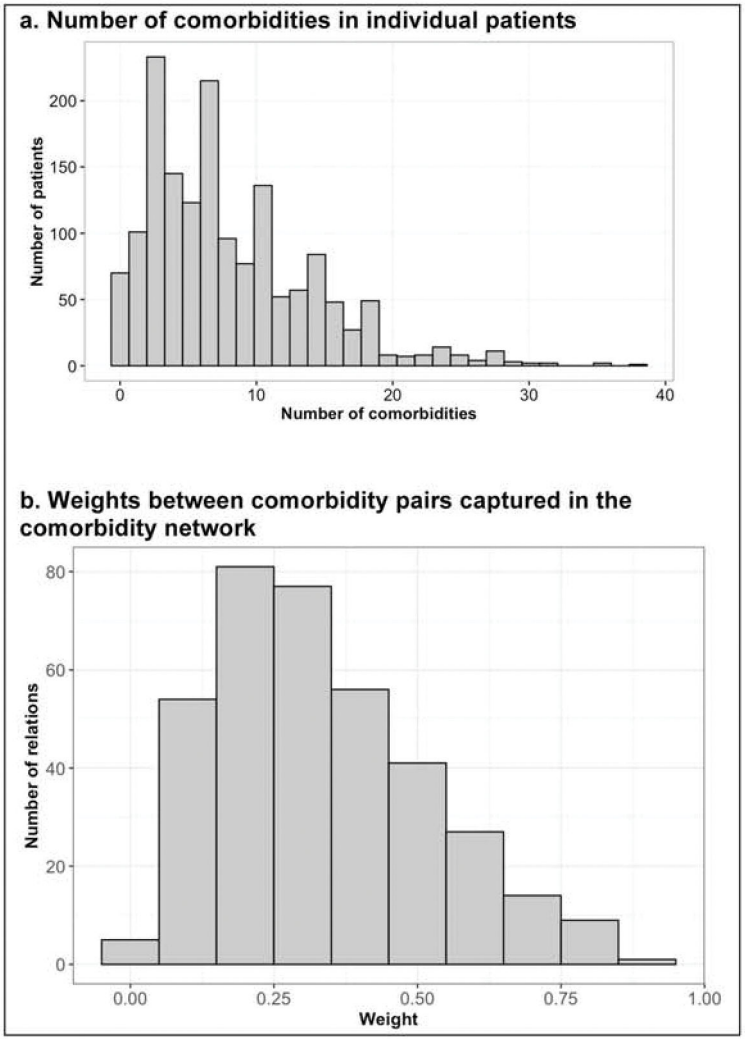
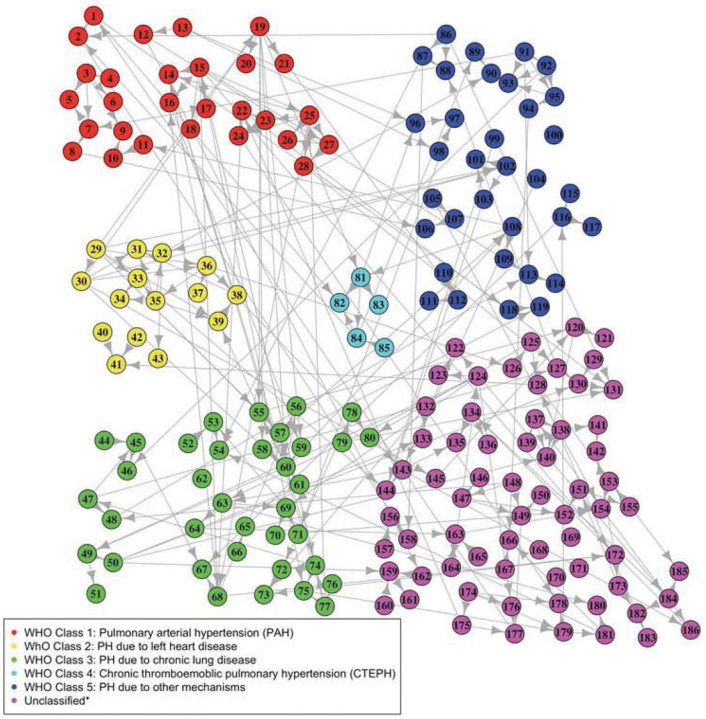


Figure 3.
Comorbidity burden in children with pulmonary hypertension



PAH
- Hereditary hemorrhagic telangiectasis (1), iron metabolism disorder (2)
- Raynaud phenomenon (3), systemic vasculitis (4), other connective tissue disease (5), systemic sclerosis (6)
- Juvenile idiopathic arthritis (7), hemophagocytic syndrome (8)
- Systemic lupus erythematosus (9), nephritis (10), essential hypertension (11)
- Portal hypertension (12), chronic liver disease (13)
- CHD left-to-right shunt (14), respiratory distress syndrome (15), emphysema (16), maternal complications of pregnancy (17), congenital pneumonia (18)
- Pleurisy (19), pericardium disorder (20), noninfectious disorders of lymphatic channels (21)
- PPHN (22), tachypnea (23), neonatal endocrine and metabolic disorder (24), exceptionally large baby (25)

PH due to left heart disease
- Congestive heart failure (26), cardiomegaly (27)
- Valvular heart disease (tricuspid or pulmonary valve disorders) (28), left heart failure (29)
- Valvular heart disease (mitral or aortic valve disorders) (30), rheumatic heart disease (31), left-sided CHD (32)
- Other CHD (33), congenital anomalies of peripheral vascular system (34)
- Conduction disorder (35), cardiac dysrhythmias (36)
- CHD right-to-left shunt (37), von Willebrand's disease (38), acute cor pulmonale (39)
- Myocarditis (40)

PH due to chronic lung disease and/or hypoxia
- Bronchiectasis (41), cystic fibrosis (42), pneumocystosis (43)
- Interstitial lung disease (44), chronic bronchitis (45)
- Pneumonia (46), pneumonitis (47), malnutrition (48)
- Pulmonary eosinophilia (49), edema of larynx (50), cerebrovascular disorder (51)
- Sleep apnea (52), intellectual disability (53)
- Other congenital anomalies of respiratory system (54), cerebral depression coma (55)
- Congenital cystic lung (56), primary atelectasis (57)
- Congenital anomalies of larynx or trachea (58), diseases of vocal cord (59), stricture and stenosis of esophagus (60)
- Congenital lung agenesis/hypoplasia (61), acquired hypertrophic pyloric stenosis (62)
- Bronchopulmonary dysplasia (63), fetal and neonatal hemorrhage (64), cerebral cyst (65), neonatal hematological disorder (66)
- Congenital anomalies of diaphragm (67), diaphragmatic hernia (68), diaphragm paralysis (69)
- Intrauterine hypoxia and birth asphyxia (70), convulsion in newborn (71), birth trauma (72)
- Pulmonary insufficiency following trauma (73), hypotension (74), acute edema of lung (75), fluid overload disorder (76), pneumothorax (77), pulmonary collapse (78), nutritional deficiency (79)
- Pulmonary hemorrhage (80)

CTEPH
- Thromboembolism (81), hemorrhagic disorder due to intrinsic factor (82), compression of vein (83)
- Pulmonary embolism (84), endocarditis (85)

PH due to other mechanisms
- Sickle cell disease (86), acute chest syndrome (87), hereditary hemolytic anemia (88)
- Acquired hemolytic anemia (89), anorexia (loss of appetite) (90)
- Thrombocytopenia (91), drug-induced pancytopenia (92), aplastic anemia (93), leukocytopenia (94), neutropenia (95)
- Splenomegaly (96), fistula of stomach or duodenum (97)
- Hemolytic disease due to isoimmunization (98)
- Coagulation defect (99), iron deficiency anemia (100), leukocystosis (168)
- Multiple congenital anomalies (101), congenital intestinal atresia and stenosis (102), congenital anomalies of kidney (103), congenital malrotation of intestine (104)
- DiGeorge syndrome (105), tetralogy of Fallot (106), primary immunodeficiency (107), velo-cardio-facial syndrome (108)
- Cri-du-chat syndrome (109), other chromosomal anomalies (110), other autosomal deletion syndromes (111), rickets (112)
- Down syndrome (113), myoneural disorder (114)
- Patau syndrome (115), Edwards syndrome (116), Hirschsprung (117), magnesium metabolic disorder (118)
- Prader-Willi syndrome (119), abnormal weight gain (120), aneurysm (121)
- Cerebral palsy (122), abnormal involuntary movement (123)
- Adrenogenital disorder (124), adrenal hypofunction (125), microcephaly (126)
- Disorder of muscle, ligament and fascia (127), abnormality of gait (128), spontaneous ecchymoses (129)
- Glycogen storage disorder (130), cardiomyopathy (131), hereditary muscular dystrophy (132)
- Plasma protein metabolic disorder (133), gangrene (134), defibrination syndrome (135)
- Acute kidney failure (136), phosphorus metabolism disorder (137)
- Hepatomegaly (138), hydrops fetalis (139), edema (140)
- Mediastinitis (141), transient mental disorder (142)
- Congenital spleen anomaly (143), situs inversus (144)
- Epilepsy (145), anoxic brain damage (146)
- Intracranial hemorrhage (147), hydrocephalus (148)
- Choanal atresia (149), congenital anomalies of skull and face (150)
- Hearing loss (151), congenital musculoskeletal deformities (152), delay in development (153)
- Congenital brain reduction deformities (154), eosinophilia (155), phlebitis and thrombophlebitis (156)
- Periventricular leukomalacia (157), pyogenic granuloma (158)
- Intestinal obstruction (159), intestinal vascular insufficiency (160)
- Hypothyroidism (161)
- Perinatal digestive system disorders (162), intestinal malabsorption (163)
- Disorder of eye movement (164), lack of coordination (165)
- Gastrointestinal hemorrhage (166), gastroenteritis and colitis (167)
- Encephalocele (169), cerebral compression (170), bilirubin excretion disorder (171), hepatitis (172)
- Disturbance of salivary secretion (173), speech disturbance (174), childbirth complications (175)
- Mixed acid base balance disorder (176), calculus of kidney (177), gastroparesis (178)
- Gonadal dysgenesis (179), drug withdrawal syndrome (180)

Unclassified (comorbidities that do not belong to a cluster and do not fit into a WHO subtype: Dyspepsia (181), diabetic mother syndrome (182), umbilical hernia (183), persistent vomiting (184), essential hypertension (185), cyanosis (186)

Figure 4.
Bayesian comorbidity network of children with pulmonary hypertension

Table 1

Mapping of network-derived comorbidity clusters with PH subtypes defined in the WHO classification – a clinical classification of PH based on similarities in pathophysiologic mechanisms, clinical presentations and therapeutic approaches.

WHO Classification of PH	Network-derived Comorbidity Clusters
1. Pulmonary arterial hypertension (PAH)	
1.1 Idiopathic PAH	• N/A
1.2 Heritable PAH	• Hereditary hemorrhagic telangiectasis, iron metabolism disorder
1.3 Drug and toxin induced	• N/A (exposures to drugs and toxins were not captured in the dataset)
1.4.1 Associated with CTD	• Raynaud's phenomenon, systemic sclerosis, systemic vasculitis, other CTD • Juvenile idiopathic arthritis, hemophagocytic syndrome • Systemic lupus erythematosus, nephritis, chronic kidney disease
1.4.2 Associated with HIV infection	N/A (there were no PH patients with HIV in the dataset)
1.4.3 Associated with portal hypertension	• Portal hypertension, chronic liver disease
1.4.4 Associated with CHD	• CHD (left to right shunt), maternal complications of pregnancy, respiratory distress syndrome, emphysema, congenital pneumonia
1.4.5 Associated with schistosomiasis	N/A (there were no PH patients with schistosomiasis in the dataset)
1′ PVOD	• Pleurisy, noninfectious disorders of lymphatic channels, pericardium disorder
1″ PPHN	• PPHN, tachypnea of newborn, neonatal endocrine and metabolic disturbance, exceptionally large baby • Intrauterine hypoxia and birth asphyxia, birth trauma, convulsion in newborn
2. PH due to left heart disease	
2.1 Left ventricular systolic dysfunction	• Congestive heart failure, cardiomegaly
2.2 Left ventricular diastolic dysfunction	
2.3 Valvular disease	• Valvular heart disease (tricuspid or pulmonary valve disorders), left heart failure • Valvular heart disease (mitral or aortic valve disorders), rheumatic heart disease, left-sided CHD
2.4 Congenital/acquired left heart disorder	• Other CHD, congenital anomalies of peripheral vascular system • Conduction disorder, cardiac dysrhythmias • CHD (RL shunt), Von Willebrand's disease, acute cor pulmonale • Myocarditis
3. PH due to lung diseases and/or hypoxia	
3.1 Chronic obstructive pulmonary disease	• Bronchiectasis, pneumocystosis, cystic fibrosis
3.2 Interstitial lung disease	• Interstitial lung disease, chronic bronchitis
3.3 Other mixed restrictive or obstructive pulmonary disease	• Pneumonia, pneumonitis, malnutrition

WHO Classification of PH	Network-derived Comorbidity Clusters
	• Pulmonary eosinophilia, edema of larynx, cerebrovascular disorder • Pulmonary insufficiency following trauma, acute edema of lung, hypotension, fluid overload disorder, pneumothorax, pulmonary collapse, nutritional deficiency • Pulmonary hemorrhage
3.4 Sleep-disordered breathing	• Sleep apnea, intellectual disability
3.5 Alveolar hypoventilation disorders	• Congenital anomalies of respiratory system, cerebral depression coma
3.6 Chronic exposure to high altitude	N/A (there are no diagnosis codes for high altitude PH)
3.7 Developmental lung diseases	• Congenital cystic lung, primary atelectasis • Congenital anomalies of larynx/trachea, diseases of vocal cord, stricture/stenosis of esophagus • Congenital lung agenesis/hypoplasia/dysplasia, acquired hypertrophic pyloric stenosis • Bronchopulmonary dysplasia, fetal and neonatal hemorrhage, cerebral cyst, neonatal hematological disorder • Congenital anomalies of diaphragm, diaphragmatic hernia, diaphragm paralysis
4. Chronic thromboembolic PH (CTEPH)	• Thromboembolism, compression of vein, hemorrhagic disorder due to intrinsic factor • Pulmonary embolism, endocarditis
5. PH due to unclear multifactorial mechanisms	
5.1. Hematological disorders	• Sickle cell disease, acute chest syndrome, hereditary hemolytic anemia • Acquired hemolytic anemia, anorexia (loss of appetite) • Thrombocytopenia, aplastic anemia, leukocytopenia, neutropenia, drug-induced pancytopenia • Coagulation defect, iron deficiency anemia • Hemolytic disease due to isoimmunization
5.2. Systemic disorders	• Essential hypertension
5.3. Metabolic disorders	• Glycogen storage disorder, hereditary muscular dystrophy, cardiomyopathy • Hypothyroidism • Plasma protein metabolic disorder, defibrination syndrome, gangrene • Mixed acid base balance disorder, calculus of kidney, gastroparesis
5.1. Others	• Acute kidney failure, phosphorus metabolism disorder • Hepatomegaly, edema, hydrops fetalis • Splenomegaly, fistula of stomach or duodenum • Mediastinitis, transient mental disorder

Table 2

Mapping of network-derived comorbidity clusters with PH subtypes defined in the Panama classification – a clinical classification of PH focusing on the causative factors in pediatric PH.

Panama Classification of PH	Network-derived Comorbidity Clusters
1. Prenatal or developmental pulmonary hypertensive vascular disease	• Congenital anomalies of diaphragm, diaphragmatic hernia, diaphragm paralysis • Congenital lung agenesis/hypoplasia/dysplasia, acquired hypertrophic pyloric stenosis • Congenital cystic lung, primary atelectasis • Congenital anomalies of respiratory system, cerebral depression coma, childbirth complications
2. Perinatal pulmonary vascular maladaptation	• PPHN, tachypnea of newborn, neonatal endocrine and metabolic disturbance, exceptionally large baby • Intrauterine hypoxia and birth asphyxia, birth trauma, convulsion in newborn
3. Pediatric cardiovascular disease	• CHD (left to right shunt), maternal complications of pregnancy, respiratory distress syndrome, emphysema, congenital pneumonia • Congestive heart failure, cardiomegaly • Valvular heart disease (tricuspid or pulmonary valve disorders), left heart failure • Valvular heart disease (mitral or aortic valve disorders), rheumatic heart disease, left-sided CHD • Other CHD, congenital anomalies of peripheral vascular system • Conduction disorder, cardi0c dysrhythmias • CHD (RL shunt), Von Willebrand's disease, acute cor pulmonale • Myocarditis
4. Bronchopulmonary dysplasia	• Bronchopulmonary dysplasia, fetal and neonatal hemorrhage, cerebral cyst, neonatal hematological disorder
5. Isolated pediatric pulmonary hypertensive vascular disease (PPHVD) or Isolated pediatric PAH	• Hereditary hemorrhagic telangiectasis, iron metabolism disorder
6. Multifactorial pulmonary hypertensive vascular disease in congenital malformation syndromes	• Down syndrome, myoneural disorder • DiGeorge syndrome, velo cardio facial syndrome, primary immunodeficiency, Tetralogy of Fallot • Edwards syndrome, Patau syndrome, Hirschsprung, magnesium metabolism disorder • Cri-du-chat syndrome, other chromosomal anomalies, other autosomal deletion syndromes, rickets • Turner syndrome (gonadal dysgenesis), drug withdrawal syndrome • Congenital anomalies of skull and face, choanal atresia • Congenital anomalies of larynx/trachea, diseases of vocal cord, stricture/stenosis of esophagus • Multiple congenital anomalies, congenital intestinal atresia and stenosis, congenital anomalies of kidney, congenital malrotation of intestine • Congenital musculoskeletal deformities, delay in development, hearing loss • Adrenogenital disorder, adrenal hypofunction, microcephaly [Smith-Lemli-Opitz syndrome] • Disorder of muscle ligament and fascia abnormality of gait, spontaneous ecchymoses [Ehlers-Danlos syndrome]*

Panama Classification of PH	Network-derived Comorbidity Clusters
	• Situs inversus, congenital spleen anomaly
7. Pediatric lung disease	• Interstitial lung disease, chronic bronchitis • Bronchiectasis, pneumocystosis, cystic fibrosis • Pneumonia, pneumonitis, malnutrition • Pulmonary eosinophilia, edema of larynx, cerebrovascular disorder • Sleep apnea, intellectual disability
8. Pediatric thromboembolic disease	• Thromboembolism, compression of vein, hemorrhagic disorder due to intrinsic factor • Pulmonary embolism, endocarditis
9. Pediatric hypobaric hypoxic exposure	• N/A (there are no diagnosis codes for hypobaric hypoxic exposure)
10. Pediatric pulmonary vascular disease associated with other system disorders	• Portal hypertension, chronic liver disease • Raynaud's phenomenon, systemic sclerosis, systemic vasculitis, other CTD • Juvenile idiopathic arthritis, hemophagocytic syndrome • Systemic lupus erythematosus, nephritis, chronic kidney disease • Sickle cell disease, acute chest syndrome, hereditary hemolytic anemia • Acquired hemolytic anemia, anorexia (loss of appetite) • Thrombocytopenia, aplastic anemia, leukocytopenia, neutropenia, drug-induced pancytopenia • Coagulation defect, iron deficiency anemia, leukocytosis • Splenomegaly, fistula of stomach or duodenum • Glycogen storage disorder, hereditary muscular dystrophy, cardiomyopathy • Hepatomegaly, edema, hydrops fetalis, edema • Splenomegaly, fistula of stomach or duodenum • Pleurisy, noninfectious disorders of lymphatic channels, pericardium disorder

Table 3

Rare subtypes identified through network analysis

Network-derived Comorbidity Clusters	Published Evidence Supporting Subtype Validity
Juvenile idiopathic arthritis (JIA) Hemophagocytic syndrome	The co-occurrence of PAH in patients with systemic-onset JIA and macrophage activation syndrome has been documented in several case series [31–32].
Glycogen storage disorder (GSD), hereditary muscular dystrophy, cardiomyopathy	This cluster captures the characteristic features of Type II GSD. PH in Type II GSD has been reported in several case studies [34–35].
Situs inversus, congenital spleen anomaly	This cluster captures the characteristic features of heterotaxy syndrome. PH associated with heterotaxy syndrome has been documented in several reports [36–37].
Cri-du-chat syndrome, other chromosomal anomalies, other autosomal deletion syndromes, rickets	Case reports have documented the co-occurrence of PH in children with Cri-du-chat, caused by underlying congenital heart disease. [38–39].
Turner syndrome (gonadal dysgenesis), drug withdrawal syndrome	Severe PH in the setting of cardiovascular abnormalities had been reported in children with Turner syndrome [40–41].
Prader Willi syndrome, abnormal weight gain, aneurysm	A case report documented a sudden death secondary to PH in a child with Prader-Willi syndrome [42]. Obstructive sleep apnea and other obesity-related comorbidities may have contributed to the development of PH.
Adrenogenital disorder, adrenal hypofunction, microcephaly	This cluster may represent Smith-Lemli-Opitz syndrome (SLOS). The co-occurrence of SLOS and PPHN/PAH has been documented in patients with altered expression of caveolin-1 [43–44].
Disorder of muscle ligament and fascia, abnormality of gait, spontaneous ecchymoses	This cluster contains symptoms suggestive of Ehlers Danlos syndrome (EDS). Spontaneous vascular or visceral rupture is a characteristic feature of vascular EDS, a subtype of EDS caused by mutations in the COL3A1 gene. The co-occurrence of EDS and mild PH has been documented in two patients [46].
Perinatal digestive system disorders, intestinal malabsorption	Severe vitamin deficiencies in the setting of pulmonary complications have been documented in several patients with PH, whereby repletion of these vitamins led to the resolution of PH [47–50].

Table 4

Unclassified network-derived comorbidity clusters associated with neurological conditions.

- Encephalocele, cerebral compression, bilirubin excretion disorder, hepatitis
- Epilepsy, anoxic brain damage
- Intracranial hemorrhage, hydrocephalus
- Congenital brain reduction deformities, eosinophilia, phlebitis and thrombophlebitis
- Periventricular leukomalacia, pyogenic granuloma
- Intestinal obstruction, intestinal vascular insufficiency
- Perinatal digestive system disorders, intestinal malabsorption
- Gastrointestinal hemorrhage, gastroenteritis and colitis
- Disorder of eye movement, lack of coordination
- Disturbance of salivary secretion, speech disturbance, childbirth complications
- Mixed acid base balance disorder, calculus of kidney, gastroparesis
- Disorder of muscle, ligament and fascia, abnormality of gait, spontaneous ecchymoses