Computer Methods and Programs in Biomedicine 165 (2018) 107–116

Contents lists available at ScienceDirect

Computer Methods and Programs in Biomedicine

journal homepage: www.elsevier.com/locate/cmpb

A novel, data-driven conceptualization for critical left heart obstruction

James M. Meza[a], Martijn Slieker[b], Eugene H. Blackstone[c], Luc Mertens[d], William M. DeCampli[e], James K. Kirklin[f], Mohsen Karimi[g], Pirooz Eghtesady[h], Kamal Pourmoghadam[e], Richard W. Kim[i], Phillip T. Burch[j], Marshall L. Jacobs[k], Tara Karamlou[l], Brian W. McCrindle[d,e,m,*], on behalf of the Congenital Heart Surgeons' Society

[a] Division of Cardiovascular Surgery, The Hospital for Sick Children, Toronto, CA
[b] Division of Pediatric Cardiology, Radboud University Medical Center, Nijmegan, the Netherlands
[c] Division of Cardiovascular and Thoracic Surgery and Department of Quantitative Health Sciences, The Cleveland Clinic, Cleveland, OH
[d] Labatt Family Heart Centre, The Hospital for Sick Children, Toronto, CA
[e] Division of Pediatric Cardiac Surgery, Arnold Palmer Children's Hospital, Orlando, FL
[f] Division of Cardiothoracic Surgery, University of Alabama-Birmingham, Birmingham, AL
[g] Division of Pediatric Cardiac Surgery, Yale-New Haven Children's Hospital, New Haven, CT
[h] Division of Cardiothoracic Surgery, St. Louis Children's Hospital, St. Louis. MO
[i] Division of Cardiothoracic Surgery, Children's Hospital of Los Angeles, Los Angeles, CA
[j] Division of Cardiothoracic Surgery, Primary Children's Medical Center, Salt Lake City, UT
[k] Division of Cardiac Surgery, Johns Hopkins Heart and Vascular Institute, Baltimore, MD
[l] Division of Thoracic and Cardiovascular Surgery, Phoenix Children's Hospital, Phoenix, AZ
[m] Department of Pediatrics, University of Toronto, Toronto, CA

ARTICLE INFO

Article history:
Received 19 June 2018
Revised 11 August 2018
Accepted 20 August 2018

Keywords:
Data science
Cluster analysis
Hypoplastic left heart syndrome
Congenital aortic stenosis

ABSTRACT

Background: Qualitative features of aortic and mitral valvar pathology have traditionally been used to classify congenital cardiac anomalies for which the left heart structures are unable to sustain adequate systemic cardiac output. We aimed to determine if novel groups of patients with greater clinical relevance could be defined within this population of patients with critical left heart obstruction (CLHO) using a data-driven approach based on both qualitative and quantitative echocardiographic measures.

Methods: An independent standardized review of recordings from pre-intervention transthoracic echocardiograms for 651 neonates with CLHO was performed. An unsupervised cluster analysis, incorporating 136 echocardiographic measures, was used to group patients with similar characteristics. Key measures differentiating the groups were then identified.

Results: Based on all measures, cluster analysis linked the 651 neonates into groups of 215 (Group 1), 338 (Group 2), and 98 (Group 3) patients. Aortic valve atresia and left ventricular (LV) end diastolic volume were identified as significant variables differentiating the groups. The median LV end diastolic area was 1.35, 0.69, and 2.47 cm^2 in Groups 1, 2, and 3, respectively ($p < 0.0001$). Aortic atresia was present in 11% (24/215), 87% (294/338), and 8% (8/98), in Groups 1, 2, and 3, respectively ($p < 0.0001$). Balloon aortic valvotomy was the first intervention for 9% (19/215), 2% (6/338), and 61% (60/98), respectively ($p < 0.0001$). For those with an initial operation, single ventricle palliation was performed in 90% (176/215), 98% (326/338), and 58% (22/38) ($p < 0.0001$). Overall mortality in each group was 27% (59/215), 41% (138/338), and 12% (12/98) ($p < 0.0001$).

Conclusions: Using a data-driven approach, we conceptualized three distinct patient groups, primarily based quantitatively on baseline LV size and qualitatively by the presence of aortic valve atresia. Man-

* Corresponding author. 555 University Ave, Room 4432, Black Wing, Toronto, ON, CA, M5G 1 × 8.
E-mail addresses: james.meza@duke.edu (J.M. Meza), brian.mccrindle@sickkids.ca (B.W. McCrindle).

https://doi.org/10.1016/j.cmpb.2018.08.014

agement strategy and overall mortality differed significantly by group. These groups roughly correspond anatomically and are analogous to multi-level LV hypoplasia, hypoplastic left heart syndrome, and critical aortic stenosis, respectively. Our analysis suggests that quantitative and qualitative assessment of left heart structures, particularly LV size and type of aortic valve pathology, may yield conceptually more internally consistent groups than a simplistic scheme limited to valvar pathology alone.

Critical left heart obstruction (CLHO), encompassing a wide spectrum of congenital heart disease from isolated aortic valve stenosis to hypoplastic left heart syndrome (HLHS), is defined as critical due to ductal-dependent circulation resulting from the left heart's inability to sustain the systemic circulation. Noonan and Nadas first described HLHS in 1958, based on analysis of 101 cardiac specimens [1]. More recently, multiple consensus statements regarding the nomenclature have been produced, culminating with the International Pediatric and Congenital Cardiac Code (IPCCC) by the Society of Thoracic Surgeons and the European Association for Cardiothoracic Surgery. For HLHS alone, the IPCCC lists 38 diagnostic categories with 29 diagnostic modifiers, and many other qualifiers (such as the status of the atrial and ventricular septa, the location and extent of left-sided hypoplasia, and more) to describe the lesion and its associated diagnoses [2,3].

This classification system, based solely on visually distinguishable characteristics, may not reflect the complex relationship between valvar morphology, ventricular anatomy, and outcomes. Apart from aortic valve atresia, it is difficult to provide prognostic information regarding surgical strategy and survival [4–6]. While the classification has remained dependent upon the qualitative assessment of gross anatomy, detailed qualitative and quantitative echocardiographic evaluation of both cardiac morphology and function has become commonplace. At the same time, the techniques available for to analyze large, detailed data sets ("big data") have rapidly increased in sophistication, allowing for the discovery of relationships not possible without large and deep data. Cluster analysis, often used in the analysis of social media, represents a fully data-driven method that has been applied infrequently in clinical analyses [7]. It allows for the exploration, characterization, and recognition of novel associations within "big data."

Given the present sophistication and ubiquity of echocardiography, the Congenital Heart Surgeons' Society (CHSS) Data Center has created an imaging core lab to define and analyze the echocardiographic spectrum of CLHO. In this study, our primary aim was to identify distinct, novel, and clinically relevant groups of patients with CLHO using only baseline quantitative and qualitative measures of cardiac morphology and function.

1. Methods

1.1. Study population

Between 2005 and 2016, 854 neonates, who were diagnosed with critical left heart obstruction and admitted to a CHSS institution at ≤ 30 days old, were enrolled from 21 institutions into a prospective inception cohort. CLHO was defined as a left-sided obstructive lesion that precluded adequate systemic output through the aortic valve. Specifically, neonates with ductal-dependent circulation with aortic valve atresia, critical aortic stenosis, hypoplastic left heart syndrome and its variants, morphologically normal but hypoplastic left-sided structures, and mitral atresia were eligible. Exclusion criteria included discordant atrioventricular or ventriculoarterial connections and first intervention at a non-CHSS institution. For this study, infants with a recording from a baseline echocardiogram, which were interpreted and underwent offline qualitative assessment at the CHSS imaging core lab, were included. Those with either missing, poor quality, or incomplete baseline echocardiograms were excluded.

1.2. Data acquisition and follow-up

Participation in the registry was voluntary and parental consent was obtained at each institution. Patient medical records, including demographic, clinical, procedural, and operative data were provided to the CHSS Data Center voluntarily and confidentially. Data were abstracted as previously described [8]. The Data Center staff contacted patients annually and subsequent medical records were obtained from enrolling institutions. Institutional review board approval was obtained at each institution.

1.3. Echocardiographic review

Baseline, pre-intervention echocardiographic recordings and reports were requested for each neonate. A single pediatric cardiologist (MGS), who was blinded to patient characteristics, clinical course, operative strategy, and mortality status, reviewed each echocardiogram according to a standardized protocol, collecting 194 qualitative and quantitative, functional and morphologic variables.

1.4. Statistical analysis

1.4.1. Variable processing and descriptive statistics

Variables with greater than 50% missing values for all patients were excluded. Based on this criterion, 136 of the 194 variables were suitable for analysis. For included variables with missing values, multiple informative imputation (five times) was performed using PROC MI [9].

The normality of continuous variables was determined using the Shapiro-Wilk test. Categorical variables are presented as frequencies and proportions and compared using the chi-square or Fisher's exact test, as appropriate. Continuous variables are presented as means with standard deviations or as medians with interquartile ranges (IQR). They were compared across groups or within groups using Student's T-test, the Mann-Whitney U test, one-way ANOVA, or Kruskal-Wallis test, as appropriate. All analyses were performed using SAS version 9.2 (SAS Institute, Inc., Cary, NC).

1.4.2. Cluster analysis

To identify unique groups of neonates using quantitative and qualitative echocardiographic measures of cardiac function and morphology, cluster analysis was employed. Cluster analysis allows for the grouping of similar patients on based on the structure of the echocardiographic data, free of any *a priori* assumptions about relationships between the variables. The clustering was determined in an unsupervised fashion, meaning that the variables were given equal weight by the clustering algorithm and that the number of clusters was not pre-determined. All variable values were first standardized to a mean of 0 and a standard deviation of 1 to account for differences in scale. No additional clinical information was entered into the cluster analysis, apart from the baseline echocardiographic variables.