



Implications of Changing Z-Score Models for Coronary Artery Dimensions in Kawasaki Disease

David L. Robinson^{1,2,3} · Adam L. Ware^{1,2} · Michael C. Sauer¹ · Richard V. Williams^{1,2} · Zhining Ou¹ · Angela P. Presson¹ · Lloyd Y. Tani^{1,2} · L. LuAnn Minich^{1,2} · Dongngan T. Truong^{1,2}

Received: 5 August 2020 / Accepted: 17 November 2020 / Published online: 4 January 2021
© Springer Science+Business Media, LLC, part of Springer Nature 2021

Abstract

Background Coronary artery abnormalities in Kawasaki disease (KD) are assessed using echocardiographic z -scores. We hypothesized that changing the coronary artery (CA) z -score model would alter diagnosis and management of children with KD.

Methods In this retrospective single-center study of children treated for KD (9/2007–1/2020), we collected echocardiographic measurements for the left anterior descending (LAD), right (RCA), and left main (LMCA) coronary arteries during 3 illness phases and calculated Boston and Pediatric Heart Network (PHN) z -scores. Agreement between Boston and PHN z -scores was assessed using Kappa (κ) and Lin's Concordance Correlation Coefficients (CCC) and Bland–Altman analysis.

Results For 904 echocardiograms from 357 children, the median Boston LAD z -score was lower than the PHN (0.3 [IQR –0.6, 1.5] vs 1.6 [IQR 0.7, 2.8], CCC 0.94 [95% CI 0.93, 0.95], moderate agreement), aggregated across all illness phases. RCA and LMCA z -scores showed substantial agreement. With conversion from Boston to PHN models, the percentage of individual LAD z -scores ≥ 2.5 increased (14.6% to 32.1%). At least one CA z -score classification changed in 213 children (59.7%) across all phases, and 48 children (13.4%) had a change that altered recommended antithrombotic strategy. Agreement between models differed by age, sex, and race.

Conclusions Conversion from Boston to PHN z -scores changed at least 1 CA z -score classification in over half of KD patients and changed recommended antithrombotic management in 13%, largely driven by LAD measurements. Since diagnosis and management of KD and KD-like diseases rely upon CA z -scores, the clinical and research implications of these findings merit further exploration.

Keywords Kawasaki disease · Coronary arteries · Z-Scores · Echocardiography

Introduction/Background

Kawasaki disease (KD) is an acute vasculitis associated with coronary artery abnormalities (CAA) and the risk of myocardial ischemia and thrombosis in affected children. In the absence of a diagnostic test, the 2017 American Heart Association (AHA) guidelines for diagnosis, risk

assessment, and management of KD rely on categories of CAA severity defined by echocardiographic z -scores during 3 illness phases: acute, subacute and convalescent [1]. Recently, studies of multisystem inflammatory syndrome in children (MIS-C) associated with COVID-19 infection have reported symptoms and coronary artery dilation and aneurysms (defined by z -scores) that mimic KD. Although the association between the two diseases remains unknown and somewhat controversial, it appears that the guidelines for KD are being incorporated into the evaluation and management of MIS-C [2, 3].

The 2017 AHA guidelines reference 7 z -score models for coronary artery dimensions and note that z -scores for an individual patient can vary between models [1]. In the USA, many echocardiography laboratories use Boston Children's Hospital's z -score model, derived from coronary artery

✉ David L. Robinson
david.robinson@hsc.utah.edu

¹ University of Utah, Salt Lake City, UT, USA

² Primary Children's Hospital, Salt Lake City, UT, USA

³ Division of Pediatric Cardiology, Department of Pediatrics, Primary Children's Outpatient Services Building, 81 N Mario Capecchi Drive, Salt Lake City, UT 84113, USA

measurements obtained from healthy children at a single center [4]. Because published nomograms of pediatric echocardiographic z -scores were limited by insufficient sample sizes to account for growth and maturation, sex and race, the Pediatric Heart Network (PHN) published a multicenter z -score model adjusted for body surface area (BSA), age, sex, race and ethnicity in 2017 [5]. As we began developing protocols for MIS-C, we sought to explore the impact of changing our practice from our usual standard of care (Boston z -scores) to the newer PHN model for evaluating coronary artery dimensions. We used coronary artery measurements of patients treated for KD, recognizing that the impact would likely be similar for patients treated for MIS-C. We hypothesized that changing the model for determining coronary artery z -scores in children with KD would alter diagnosis and management and that age, sex and race would affect agreement between models.

Methods

In this single-center retrospective study, we searched the hospital database for all patients < 18 years old who received intravenous immunoglobulin (IVIG) for KD treatment between September 2007 (when Boston z -scores were adopted as the local standard of care) and January 2020. We excluded echocardiograms performed at outside centers and patients treated at an outside institution. Since our objective was to assess the impact of changes in z -scores on real-world treatment decisions, we excluded echocardiograms with no z -scores recorded in the electronic medical record (EMR) and did not review the images or remeasure the coronary artery dimensions. The Institutional Review Boards of the University of Utah and Primary Children's Hospital approved this study under a waiver of consent.

Demographic and Clinical Data

We collected demographic data from the EMR, including age at diagnosis, sex, and self-reported race/ethnicity. We divided patients into age categories similar to those reported for the PHN Normal Echocardiogram Database but combined the older age groups (since most children with KD are < 5 years old [1]) [5], using age categories of < 1 month, 1 month to 3 years, 3–6 years, and 6–18 years in our analysis. We used the race categories defined in the PHN study: White, Black and Other (including Asian, Pacific Islander, Native American, Middle Eastern, and multiracial, and Hispanic as race or as ethnicity if there was no other racial designation) [5]. We coded patients without race/ethnicity data as “unknown”.

To categorize KD as complete or incomplete, we used AHA standard definitions [1]. Children who did not meet

clinical, laboratory, or echocardiographic criteria for either complete or incomplete KD, but had prolonged fever and were treated with IVIG for possible KD were categorized as “other”. We defined antithrombotic strategies by class I and class IIa AHA recommendations based on coronary artery z -scores: < 10.0, low dose aspirin (ASA) alone and ≥ 10 , low dose ASA and anticoagulation (warfarin or low molecular weight heparin). In the convalescent phase, continued low-dose ASA would be recommended for z -scores ≥ 2.0 and ASA could be discontinued for z -scores < 2.0 [1]. The authors (DT and DR) reviewed changes in each coronary artery z -score classification within each phase and determined whether z -score changes between the Boston and PHN models altered recommended antithrombotic strategies.

Echocardiographic Data

Echocardiographic reports were reviewed for BSA, LAD, RCA and LMCA dimensions and Boston z -scores within 3 illness phases, timed from diagnosis: acute, within 6 days; subacute, 1 week to 1 month; and convalescent, 1–3 months. No echocardiograms beyond the third month after diagnosis were included in the analysis. The clinically reported Z -scores were obtained using the Boston regression equations [6]. We calculated PHN z -scores from the absolute coronary artery diameters and BSA using the published PHN regression equations [5].

The echocardiogram with the most coronary artery data for each illness phase was selected for analysis. If several echocardiograms had complete data, the first study during the illness phase was analyzed. CAA were defined using AHA z -score classifications: no involvement (< 2); dilation only (≥ 2 to < 2.5); small aneurysm (≥ 2.5 to < 5); medium aneurysm (≥ 5 to < 10 and absolute dimension < 8 mm); giant aneurysm (≥ 10 or an absolute dimension ≥ 8 mm). For patients not meeting complete KD criteria, a z -score > 2.5 for the LAD or RCA was used as an independent criterion for diagnosing incomplete KD, consistent with AHA guidelines [1].

Statistical Analysis

Demographic and clinical data were summarized using counts and percentages for categorical variables and median and interquartile range (IQR) for continuous variables due to distribution skew. Changes in CAA classifications were enumerated at each illness phase. Proportions of change in recommended antithrombotic therapy per change in CAA classification at each phase were calculated.

We assessed the degree of agreement between Boston and PHN z -scores for each coronary artery within each of the 3 illness phases as a continuous variable using Lin's Concordance Correlation Coefficient (CCC) and 95%

confidence intervals (CI) [7, 8]. The interpretation of CCC values followed McBride's recommendations, in which the lower limit of a CCC 95% CI < 0.90 indicates poor agreement; 0.90 to < 0.95, moderate; 0.95 to 0.99, substantial; and > 0.99 indicates nearly perfect agreement [9]. We performed similar assessments with dichotomized *z*-score categories of < 2.5 and ≥ 2.5 to distinguish CAA from measurements with lower *z*-scores using unweighted Cohen's Kappa (κ) coefficients with 95% CI. The interpretation of κ values followed recommendations from Altman et al., in which the lower limit of a κ 95% CI < 0 indicates no agreement; 0.00–0.21, slight; 0.21–0.40, fair; 0.41–0.60, moderate; 0.61–0.80, substantial; 0.81–0.99, almost perfect; and 1.0 indicates perfect agreement [10]. We further compared the agreement between Boston and PHN *z*-scores for LAD, RCA, and LMCA dimensions aggregated across all illness phases and obtained estimates of CCC for repeated measures using R package "cccrm" [11]. We constructed Bland–Altman plots to further evaluate the agreement between PHN and Boston *z*-scores and determine whether the variability of the discrepancy between the 2 models increased with higher *z*-scores [12].

The agreement of *z*-score models between categories of age, sex, and race was compared using CCC for repeated measures as described above, but with an 83.7% CI. The reason for using an 83.7% CI instead of a 95% CI to assess agreement differences is that there is sampling variability associated with both estimates, and thus 95% CIs can marginally overlap even if the two estimates are significantly different ($p \leq 0.05$) [13–15]. Thus, a narrower 83.7% CI is needed to equate overlap of CIs with a lack of statistical significance ($p > 0.05$). Using CIs as opposed to *p*-values encourages a descriptive interpretation of these comparisons, as determination of an optimal procedure to adjust for multiple comparisons in this case is not straightforward. Differences in levels of agreement between demographic subgroups (e.g., male vs female) were statistically non-significant (p -value > 0.05) if their corresponding 83.7% CI overlapped. Forest plots were constructed to visually depict these data. Statistical analyses were implemented using R v. 3.6.0 [11].

Results

Study Population

Of the 357 children who met all inclusion and no exclusion criteria (Table 1), 63.9% were male, 76.8% were White and 56.6% were 1 month to 3 years of age at the time of diagnosis. No child was under 1 month of age. Criteria for complete KD were met in 66.9% of children. A total of 904 echocardiograms had sufficient data for analysis: 345 (96.6%) of

Table 1 Summary of demographic data of patients treated for Kawasaki disease

	(<i>N</i> = 357)
Age (years) at diagnosis, median and IQR	2.6 (1.2, 4.4)
<i>Age group</i>	
1 month to 3 years	202 (56.6%)
3–6 years	102 (28.6%)
6–18 years	53 (14.8%)
Male	228 (63.9%)
<i>Race</i>	
White	274 (76.8%)
Black	13 (3.6%)
Other	50 (14%)
Unknown	20 (5.6%)
<i>Type of Kawasaki disease</i>	
Complete	239 (66.9%)
Incomplete	105 (29.4%)
Other	13 (3.6%)
<i>Number of analyzed echocardiograms in the first 3 months of diagnosis, per child</i>	
3	216 (60.5%)
2	115 (32.2%)
1	26 (7.3%)

children had qualifying studies during the acute phase, 286 (80.1%) during the subacute phase, and 273 (76.5%) during the convalescent phase. Acceptable echocardiograms were available for analysis in all phases for 216 (60.5%) children. Among the 141 children without acceptable echocardiograms in at least one illness phase, 8 had studies from outside centers, 16 had studies with no Boston *z*-scores recorded, and the remainder had no follow-up at our center at an illness phase during the first 3 months after diagnosis.

Comparison of Coronary Artery Z-Scores

All Boston LAD *z*-scores were lower than the corresponding PHN *z*-scores with a median difference of 1.3 (IQR 1.3, 1.3) across all illness phases (Table 2). Comparison of aggregated data for the LAD across all illness phases showed only moderate agreement between models. In contrast, Boston and PHN *z*-scores for the RCA and LMCA had substantial agreement at all illness phases. Most Boston RCA *z*-scores and nearly all LMCA *z*-scores were within 0.5 of the PHN *z*-scores (86.5% and 98.8%, respectively).

With conversion from Boston to PHN *z*-scores, the percentage of LAD *z*-scores ≥ 2.5 more than doubled in all illness phases (Table 2). In contrast, the percentage of *z*-scores ≥ 2.5 decreased in all phases when converting from Boston to PHN *z*-scores for the RCA and LMCA but the difference was $\leq 5.2\%$ in each phase (Table 2). When

Table 2 Z-score distribution and percentage of echocardiograms with Z-scores ≥ 2.5 in Boston and PHN models

Acute phase (echocardiograms, $N=345$)	Boston	PHN	CCC/ κ (95% CI)*
LAD z-score ($N=344$) Median (IQR)	0.7 (−0.4, 2.1)	2.0 (0.9, 3.4)	0.92 (0.91, 0.93)
LAD z-score ≥ 2.5	70 (20.3%)	145 (42.2%)	0.52 (0.43, 0.60)
RCA z-score ($N=344$) Median (IQR)	0.9 (0.0, 1.7)	0.7 (0.0, 1.4)	0.97 (0.97, 0.97)
RCA z-score ≥ 2.5	54 (15.7%)	36 (10.5%)	0.77 (0.67, 0.87)
LMCA z-score ($N=342$) Median (IQR)	0.6 (−0.3, 1.3)	0.5 (−0.2, 1.2)	0.99 (0.99, 0.99)
LMCA z-score ≥ 2.5	23 (6.7%)	16 (4.7%)	0.81 (0.71, 0.91)
Subacute phase (echocardiograms, $N=286$)			
LAD z-score ($N=284$) Median (IQR)	0.3 (−0.6, 1.4)	1.6 (0.7, 2.7)	0.96 (0.95, 0.96)
LAD z-score ≥ 2.5	42 (14.8%)	85 (29.9%)	0.58 (0.47, 0.68)
RCA z-score ($N=284$) Median (IQR)	0.5 (−0.3, 1.4)	0.4 (−0.2, 1.1)	0.98 (0.97, 0.98)
RCA z-score ≥ 2.5	34 (12%)	28 (9.9%)	0.89 (0.81, 0.98)
LMCA z-score ($N=284$) Median (IQR)	0.3 (−0.4, 1.1)	0.3 (−0.3, 1.0)	0.99 (0.99, 0.99)
LMCA z-score ≥ 2.5	21 (7.4%)	17 (6%)	0.83 (0.74, 0.93)
Convalescent phase (echocardiograms, $N=273$)			
LAD z-score ($N=268$) Median (IQR)	−0.1 (−0.9, 0.9)	1.2 (0.4, 2.2)	0.96 (0.95, 0.96)
LAD z-score ≥ 2.5	19 (7.1%)	58 (21.6%)	0.43 (0.3, 0.57)
RCA z-score ($N=271$) Median (IQR)	0.2 (−0.5, 1.1)	0.2 (−0.3, 0.8)	0.97 (0.97, 0.97)
RCA z-score ≥ 2.5	20 (7.4%)	15 (5.5%)	0.85 (0.72, 0.98)
LMCA z-score ($N=271$) Median (IQR)	0.1 (−0.7, 0.8)	0.1 (−0.6, 0.7)	0.99 (0.99, 0.99)
LMCA z-score ≥ 2.5	11 (4.1%)	11 (4.1%)	0.81 (0.68, 0.94)
All illness phases (total echocardiograms, $N=904$)			
LAD z-score ($N=896$) Median (IQR)	0.3 (−0.6, 1.5)	1.6 (0.7, 2.8)	0.94 (0.93, 0.95)
LAD z-score ≥ 2.5	131 (14.6%)	288 (32.1%)	0.53 (0.46, 0.60)
RCA z-score ($N=899$) Median (IQR)	0.5 (−0.3, 1.5)	0.4 (−0.2, 1.1)	0.97 (0.97, 0.98)
RCA z-score ≥ 2.5	108 (12.0%)	79 (8.8%)	0.83 (0.77, 0.89)
LMCA z-score ($N=897$) Median (IQR)	0.3 (−0.4, 1.1)	0.3 (−0.4, 1.0)	0.99 (0.99, 0.99)
LMCA z-score ≥ 2.5	55 (6.1%)	44 (4.9%)	0.82 (0.73, 0.90)

*Lin's Concordance Correlation Coefficient (CCC) is used for continuous variables, κ for dichotomous variables. Interpretation of CCC: CCC 95% CI with lower limit value: <0.90 =poor agreement; 0.90 to <0.95 =moderate; 0.95 to 0.99 =substantial; >0.99 =nearly perfect agreement [9]. Interpretation of κ : κ 95% CI with lower limit: value <0 =no agreement; 0.00 – 0.21 =slight agreement; 0.21 – 0.40 =fair agreement; 0.41 – 0.60 =moderate agreement; 0.61 – 0.80 =substantial agreement; 0.81 – 0.99 =almost perfect agreement; 1.0 =perfect agreement [10]

assessing all 3 coronary artery dimensions across all phases, the number of individual z-scores ≥ 2.5 increased by 39.8% (294 using Boston vs 411 using PHN z-scores).

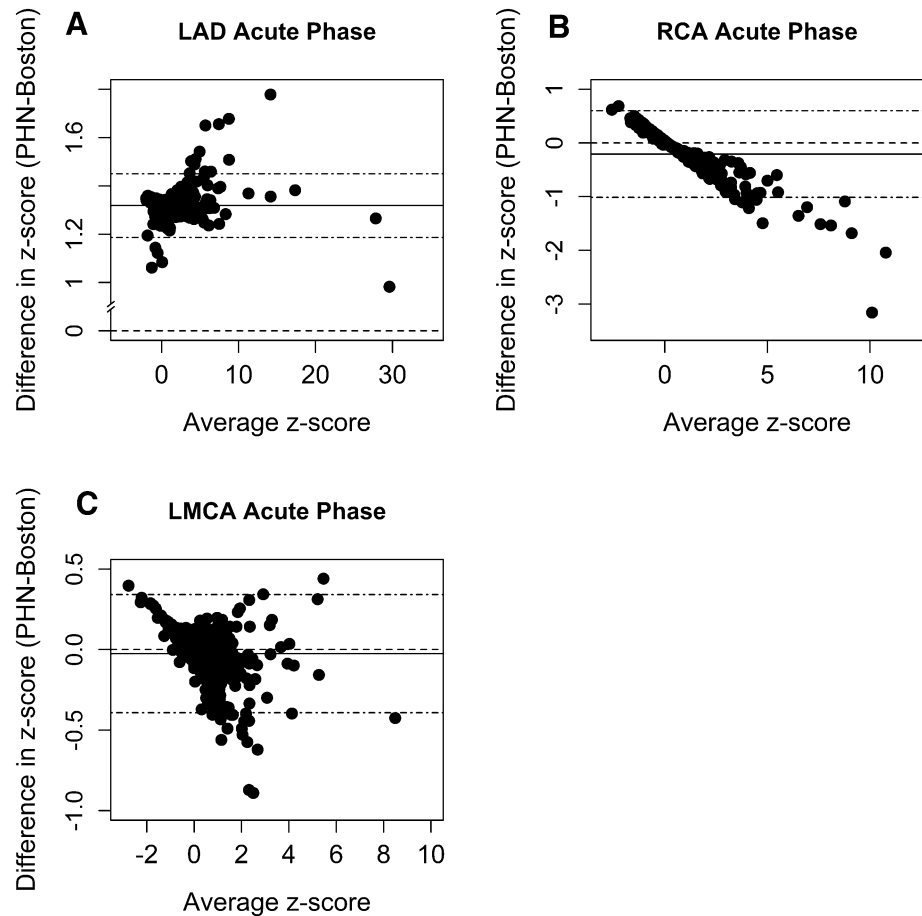
The consistency of the difference between the Boston and PHN LAD z-scores was only slightly impacted by z-score magnitude when the average of paired z-scores exceeded 5 (Fig. 1). Since Bland–Altman plots for each illness phase were consistent for each coronary artery, only the acute phase plots are shown.

Over half (59.7%, 213/357) of the study population had at least one change in coronary artery z-score classification when converting between models. When changes in LAD z-score classification were excluded, this percentage decreased to 21.2% (76/357). The LAD accounted for 72.0% (265/368) of all individual z-score classification

changes, with 16.1% (144/896) of all LAD measurements changing by 1 classification level, and 13.5% (121/896) changing by 2 levels. Changes most frequently occurred from the no involvement/dilation only level to the small aneurysm level (Table 3). For the RCA, 8.3% (75/899) changed by 1 level and 0.6% (5/899) changed by 2 levels. All 5 children who had giant RCA aneurysms by the Boston model were reclassified as having moderate-sized aneurysms by the PHN model. For the LMCA, only 3.5% (31/897) changed by 1 level of classification, and 4 (0.4%) changed by 2 levels.

Of the 13 children originally classified as “other” KD using the Boston model, 7 children (54%) were reclassified as meeting criteria for incomplete KD with the PHN model, given an increase in their LAD z-score to ≥ 2.5 .

Fig. 1 Bland–Altman plots for each coronary dimension in the acute phase of illness. X-axis represents the subject-specific average between two z-scores (PHN z-score + Boston z-score/2). Y-axis represents the subject-specific difference of the two z-scores (PHN z-score – Boston z-score). The dashed line intersecting zero on the y-axis indicates the reference value of no discrepancy/bias. The solid horizontal line indicates the mean of the differences. The farther the solid line (average difference) departs from zero, the larger the overall discrepancy between the two z-score systems. Stippled (“dot-dash”) lines bound the 95% CI of mean differences



Impact of Differences in Z-Score Models on Antithrombotic Recommendations

Of the 213 children with a change in AHA coronary artery z-score classification, recommendations for antithrombotic management changed in at least one illness phase for 22.5% (48/213). Antithrombotic therapy would change for 1.1% (2/175) of all individual z-score reclassifications during the acute phase, 5.4% (6/112) in the subacute phase, and 76.3% (71/93) in the convalescent phase (Table 4). Most of the changes in the convalescent phase (55/71, 77.5%) were related to the recommendation to continue low dose ASA beyond that phase, largely driven by increases in LAD z-scores (54/55, 98%).

Effects of Age, Sex and Race on the Level of Agreement Between Models

There was substantial agreement between Boston and PHN LAD scores in the 3–6 years age category, but only moderate agreement in the 1 month to 3 years category and poor agreement in the 6–18 years category. For the RCA, substantial agreement was found in the 1 month to 3 years category, compared to moderate agreement in the 6–18 years category. All

age groups had substantial agreement for the LMCA (Fig. 2a). Males demonstrated substantial agreement between Boston and PHN LAD z-scores while females had poor agreement. There was no difference in agreement based on sex for the RCA or LMCA z-scores (Fig. 2b). White children showed moderate agreement in LAD z-scores between models, while children of non-White races had poor agreement. There was no significant difference in RCA or LMCA z-score agreement by race (Fig. 2c).

The proportions of changes in AHA z-score classification after conversion from Boston to PHN z-scores varied significantly by age and sex, but not by race. Of 102 children aged 3–6 years, 41.2% (42) had a change in at least one coronary artery z-score categorization in at least one phase of illness, compared to 67.3% (136/202) of those 1 month to 3 years of age, and 66% (35/53) of those 6–18 years of age, ($p < 0.001$). Of 129 female patients, 48.8% (63) had a change in z-score categorization compared to 65.8% (150/228) of males ($p = 0.002$).

Table 3 Summary of changes in Z-score classification from Boston to PHN models for all echocardiograms (*N*)

Change in z-score classification ^a (Boston → PHN)	Acute (<i>N</i> =345)	Subacute (<i>N</i> =286)	Convalescent (<i>N</i> =273)	All phases (<i>N</i> =904)
<i>LAD</i>	(<i>N</i> =344)*	(<i>N</i> =284)	(<i>N</i> =268)	(<i>N</i> =896)
No change	219 (63.7%)	210 (74.0%)	202 (75.4%)	631 (70.4%)
No involvement → dilation only	30 (8.7%)	24 (8.5%)	23 (8.6%)	77 (8.6%)
No involvement → small aneurysm	55 (16.0%)	35 (12.3%)	31 (11.6%)	121 (13.5%)
Dilation only → small aneurysm	20 (5.8%)	8 (2.8%)	8 (2.9%)	36 (4.0%)
Small → medium aneurysm	20 (5.8%)	5 (1.7%)	3 (1.1%)	28 (3.1%)
Medium → giant aneurysm	–	2 (0.7%)	1 (0.4%)	3 (0.3%)
<i>RCA</i>	(<i>N</i> =344)	(<i>N</i> =284)	(<i>N</i> =271)	(<i>N</i> =899)
No change	307 (89.2%)	259 (91.2%)	253 (93.4%)	819 (91.1%)
Dilation only → no involvement	14 (4.1%)	9 (3.2%)	8 (3.0%)	31 (3.5%)
Small aneurysm → no involvement	2 (0.6%)	2 (0.7%)	1 (0.3%)	5 (0.6%)
Small aneurysm → dilation only	16 (4.6%)	4 (1.4%)	4 (1.5%)	24 (2.7%)
Medium → small aneurysm	3 (0.9%)	6 (2.1%)	4 (1.5%)	13 (1.4%)
Giant → medium aneurysm	2 (0.6%)	4 (1.4%)	1 (0.3%)	7 (0.7%)
<i>LMCA</i>	(<i>N</i> =342)	(<i>N</i> =284)	(<i>N</i> =271)	(<i>N</i> =897)
No change	329 (96.2%)	271 (95.4%)	262 (96.6%)	862 (96.1%)
No involvement → dilation only	1 (0.3%)	2 (0.7%)	1 (0.4%)	4 (0.4%)
Dilation only → no involvement	5 (1.4%)	5 (1.7%)	4 (1.5%)	14 (1.6%)
Dilation only → small aneurysm	–	1 (0.4%)	2 (0.7%)	3 (0.3%)
Small aneurysm → no involvement	2 (0.6%)	1 (0.4%)	1 (0.4%)	4 (0.4%)
Small aneurysm → dilation only	5 (1.4%)	4 (1.4%)	1 (0.4%)	10 (1.1%)

^aZ-score classification: No involvement: <2; dilation only: ≥2 to <2.5; small aneurysm: ≥2.5 to <5; medium aneurysm: ≥5 to <10 and absolute dimension <8 mm; giant aneurysm: >10 or absolute dimension ≥8 mm [1]

*Italicized “*N*” headings denote total number of available coronary artery measurements at each location for each phase

Table 4 Summary of changes in recommended antithrombotic strategy due to change in AHA Z-score classification with conversion from Boston to PHN models

Change in antithrombotic strategy ^a (Boston → PHN)	Acute* (<i>N</i> =175)	Subacute (<i>N</i> =112)	Convalescent (<i>N</i> =93)	All phases (<i>N</i> =380)
No change	173 (98.9%)	106 (94.6%)	22 (23.7%)	301 (79.2%)
<i>Increase intensity</i>				
No anticoagulation → anticoagulation recommended	–	2 (1.8%)	1 (1.1%)	3 (0.8%)
Stop ASA → continue ASA	–	–	55 (59.1%)	55 (14.5%)
<i>Decrease intensity</i>				
Anticoagulation → no anticoagulation	2 (1.1%)	4 (3.6%)	1 (1.1%)	7 (1.8%)
Continue ASA → stop ASA	–	–	14 (15.0%)	14 (3.7%)

^aAntithrombotic Strategies- Coronary artery z-scores ≥10 treated with anticoagulant (warfarin or low molecular weight heparin) and ASA. In the convalescent phase, z-scores ≥2.0 would continue ASA, whereas ASA can be stopped if <2.0 [1]

**N* refers to total number of z-scores reclassified with conversion from Boston to PHN models at each phase

Discussion

Although previous studies have explored differences between z-score models [16–19], we investigated the actual impact of these differences on KD diagnosis and management. Our findings are particularly relevant now,

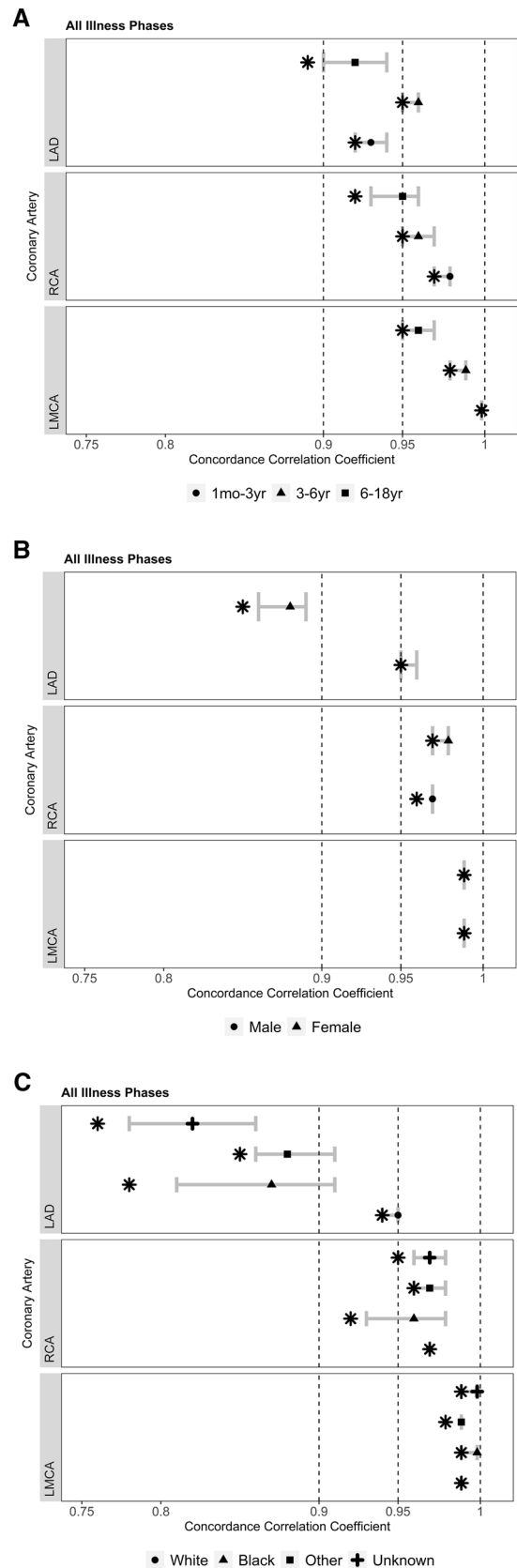
given the emergence of the KD-like syndrome known as MIS-C that has been newly described during the COVID-19 pandemic [2, 3]. Since KD and MIS-C share similar features and clinicians are referring to KD guidelines in the management of MIS-C, the methods of determining coronary artery size normalized to body size in growing children have drawn increasing attention. We identified

Fig. 2 Forest plots depict agreement between the Boston and PHN z -score models compared by: **a** age, **b** sex, and **c** race. Symbols in the keys represent demographic subgroups, with the location of each symbol along the 83.7% CI denoting CCC point estimates. Differences in levels of agreement between demographic subgroups (e.g., male vs female) were statistically non-significant (p -value > 0.05) if their corresponding 83.7% CI overlapped [13–15]. Asterisks represent the lower limit of the 95% CI, which corresponds to the level of agreement between the Boston and PHN models: < 0.90 indicates poor agreement; 0.90 to < 0.95, moderate; 0.95 to 0.99, substantial; and > 0.99 indicates nearly perfect agreement [9]

important effects on diagnosis of CAA and recommended antithrombotic therapy when converting from the commonly used Boston z -score model to the newly published PHN model. Conversion from Boston to PHN z -scores yielded at least one change in CA z -score classification for nearly 60% of children and changed recommended antithrombotic strategy in 13.4%. PHN LAD z -scores were always higher than Boston z -scores, showing the weakest correlations between the models (compared to the RCA and LMCA) and accounting for 72.0% of all changes in CAA classification. It is important to note that nearly 14% of all LAD measurements increased by 2 levels of z -score classification. Overall, we found a nearly 40% increase in CAA with z -score ≥ 2.5 . These findings illustrate the practical implications of using different models in the evaluation of a disease for which the standard of care is based on z -scores.

Despite some methodological differences, our results are supported by prior reports. When the PHN investigators compared Boston and PHN z -scores of commonly measured cardiac structures (abstract), one of the poorest correlations between the two models was for LAD z -scores, with a median difference of 1.33 (PHN-Boston) [16]. Similarly, Lorenzoni, et al. (abstract) compared Boston and PHN coronary artery z -scores for local KD patients and those enrolled in the PHN Randomized Trial of Pulse Steroid Therapy for Primary Treatment of KD and reported a 50% increase in abnormalities as well as a change in the AHA recommended antithrombotic strategy in 21.5% of patients using PHN z -scores, defining CAA as a z -score of 2.0 or higher [17]. Similar to our study, the higher frequency of CAA using the PHN model was driven by higher LAD z -scores, and AHA antithrombotic treatment category changed for a relatively large number of patients. However, we also found that over 50% of patients treated for KD who did not meet criteria for complete or incomplete KD using Boston z -scores had PHN LAD z -scores ≥ 2.5 , satisfying the AHA diagnostic criteria for diagnosing incomplete KD. Unfortunately, we have no data regarding the coronary artery z -scores of children with similar clinical findings who were not treated for KD.

Disagreements between the Boston model and other z -score models have also been reported [5, 18]. In contrast



to our findings regarding the PHN model, Ronai, et al., found Boston coronary artery LAD z -scores tended to be higher when compared to Washington D.C. and Quebec models. These investigators plotted coronary artery dimensions corresponding with a z -score of 10.0 in each model vs. BSA and found the dimensions of both the LAD and RCA for the Boston model were significantly lower than the Washington D.C. model and similar to the Quebec model. Agreement between models was poorest for measurements with the highest LAD or RCA Boston z -scores (range 7–14): 50% had Boston z -scores ≥ 10 while none of the corresponding D.C. and only 30% of Quebec z -scores were ≥ 10 for the same dimensions [19]. In our study, for larger coronary artery dimensions (when the average of paired z -scores exceeded 1), the Boston RCA z -scores were indeed higher than the PHN, and all 5 patients with RCA giant aneurysms using Boston z -scores had only moderate-sized aneurysms using PHN z -scores (despite overall substantial agreement between the Boston and PHN RCA z -scores). Conversely, our study showed that Boston LAD z -scores were consistently lower than the PHN, with increasing variation in the magnitude of that difference when the average of paired z -scores exceeded 5.

The PHN Normal Echocardiogram database study found that age, sex, and race contributed little to the variation in the z -score models for coronary artery dimensions [5]. As the first study to examine the effects of age, sex, and race on agreement between pediatric coronary artery z -score models in a real-world setting, we found that these factors had an impact on the level of agreement between Boston and PHN z -score models, with higher proportions of males and children in the extreme age groups (1 month to 3 years and 6–18 years) having at least one change in classification. While children of non-White races had poorer agreement between LAD z -scores than White children, the proportions of changes in AHA z -score classification did not significantly differ by race. These differences may be due to underrepresentation of certain groups in single-center z -score models, such as non-White race groups, leading to lower levels of agreement when compared to a system derived from a more evenly diverse population. While the Boston model provided no breakdown of age, sex, or race [4], the PHN model was specifically designed for equal representation of various races (White, Black, and Other), sexes, and ages [5]. Therefore, the demographics of each z -score model population may need to be more carefully examined to determine the generalizability to the local population served by each center.

The reason for these differences in z -scores between the Boston and PHN models is likely multifactorial, including differences in sample size, definitions of normal or healthy children, proportions of ages, sexes, or races of patients comprising the study cohorts, methods of normalizing raw

dimensions, and number of observers making measurements, as well as changes in technology and image resolution. For example, Boston z -score regression equations were derived from measurements by multiple observers at a single center obtained between 1987 and 2000 from 221 healthy children who were predominantly White [4]. In communication with the authors, it was confirmed that new patients have not been added to the Boston database for coronary artery dimensions. PHN coronary z -scores were derived from measurements by 2 core lab observers of 2507 healthy children of nearly equal racial distribution from 19 North American centers collected between 2013 and 2015 [5]. The poor correlation between models for LAD z -scores, in particular, may be attributable to greater difficulty in imaging the LAD in comparison to the more proximal LMCA and RCA. Furthermore, spatial resolution for available echocardiographic technology may lead to greater variability and random error in measurement of small structures such as the LAD. Both of these issues could have contributed to differences in the measurements used to derive regression equations for the LAD between the two models. Although we suggest potential reasons for differences among z -score models, it is important to note that our study was not designed to determine the best z -score model and we have no outcome data to evaluate the newly published PHN model. Nevertheless, we did highlight some of the pitfalls in applying two different z -score models derived from healthy children to children with a specific disease. While use of multiple z -score systems is common practice in pediatric cardiology, it is important to recognize that models are not interchangeable and the z -score model for clinical decision-making should be the same model from which outcome data driving the guidelines and recommendations are derived.

Limitations

This retrospective study had several limitations. First, measurements of coronary arteries were obtained from the EMR without reviewing the echocardiograms. Thus, reported measurements could have been influenced by a subjective impression of coronary artery size at the time of image acquisition and may have been adjusted to fit z -scores to the threshold supporting a diagnosis of KD. Regardless, the aim of this study was to understand the impact of changes to z -scores available at the time of clinical decision-making and not to validate measurements. Second, children for whom KD was considered but not ultimately treated were not captured in the database, so the effect of a change in z -score leading to a change in diagnosis of these patients could not be fully explored. Because only one echocardiogram for each phase of illness was analyzed, the largest coronary artery measurements for a given child may not have been captured; thus, the impact of change in z -scores over

time (e.g., regression from large to medium coronary artery aneurysms between the acute and convalescent phases) on antithrombotic strategy could not be evaluated. Third, although the impact of age, sex, and race was statistically significant, it is not clear that these differences are clinically significant. Based on published thresholds for larger arteries, the PHN investigators attributed measurement differences of up to 5% to measurement variability and thus, defined clinical significance as a difference of 5% or more between actual and predicted measurements [5]. Although this difference may be useful for larger arteries, there are no similar data for the coronary arteries, for which the variability is likely higher given their far smaller dimensions. Finally, we can make no recommendation regarding the best z -score model since, unlike other models [6, 20, 21], the association of z -scores and KD outcomes has not yet been reported for the PHN model.

Conclusions

Conversion from Boston to PHN z -scores resulted in differences in CAA classification, recommended antithrombotic strategies, and even in KD diagnosis. Differences were most frequently seen for LAD z -scores. These findings have implications for patients with KD and the KD-like disease, MIS-C, associated with COVID-19. Selection of a z -score system for risk stratification of CAA should include consideration of similarity of the model population to the local population being evaluated. The model used for making the outcome-based recommendations contained in the AHA guidelines should be the model used in clinical practice. Consistent application of a z -score model for longitudinal evaluation of CAA in each patient is critical. As much is still unknown about MIS-C, investigators leading research protocols may benefit from making an a priori determination of which z -score model(s) should be used to define inclusion criteria and determine the association between short and long-term outcomes. As more outcome data are collected using the PHN z -scores, we may be better able to understand the differences between models.

Acknowledgements This investigation was supported by the University of Utah Population Health Research (PHR) Foundation, with funding in part from the National Center for Research Resources and the National Center for Advancing Translational Sciences, National Institutes of Health, through Grant UL1TR002538 (formerly 5UL1TR001067-05, 8UL1TR000105 and UL1RR025764).

Author Contributions Conceptualization: DR, LM, LT and DT. Methodology: DR, LM, RW, AP, ZO and DT. Formal analysis and investigation: DR, MS, ZO, AP, and DT. Writing—original draft preparation: DR and DT. Writing—review and editing: DR, AW, MS, RW, ZO, AP, LT, LLM, and DT. Resources: LM, RW, LT and DT. Supervision: DT.

Compliance with Ethical Standards

Conflict of interest All the authors declared that they have no conflict of interest.

Ethical Approval Approved by the University of Utah IRB.

Consent to Participate Waived.

Consent for Publication Approved by the University of Utah IRB.

References

1. McCrindle BW, Rowley AH, Newburger JW et al (2017) Diagnosis, treatment, and long-term management of Kawasaki disease: a scientific statement for health professionals from the American Heart Association. *Circulation* 135:927–999. <https://doi.org/10.1161/CIR.0000000000000484>
2. Feldstein LR, Rose EB, Horwitz SM et al (2020) Multisystem inflammatory syndrome in U.S. Children and Adolescents. *N Engl J Med*. <https://doi.org/10.1056/NEJMoa2021680>
3. Whittaker E, Bamford A, Kenny J et al (2020) Clinical characteristics of 58 children with a pediatric inflammatory multisystem syndrome temporally associated with SARS-CoV-2. *JAMA*. <https://doi.org/10.1001/jama.2020.10369>
4. Sluysmans T, Colan SD (2005) Theoretical and empirical derivation of cardiovascular allometric relationships in children. *J Appl Physiol* 99:445–457. <https://doi.org/10.1152/jappphysiol.01144.2004>
5. Lopez L, Colan S, Stylianou M et al (2017) Relationship of echocardiographic Z scores adjusted for body surface area to age, sex, race, and ethnicity: the pediatric heart network normal echocardiogram database. *Circ Cardiovasc Imaging* 10(11):e006979. <https://doi.org/10.1161/CIRCIMAGING.117.006979>
6. McCrindle BW, Li JS, Minich LL, Colan SD, Atz AM, Takahashi M, Vetter VL, Gersony WM, Mitchell PD, Newburger JW (2007) Coronary artery involvement in children with Kawasaki disease: risk factors from analysis of serial normalized measurements. *Circulation* 116(2):174–179. <https://doi.org/10.1161/CIRCULATIONAHA.107.690875>
7. Morgan CJ, Aban I (2016) Methods for evaluating the agreement between diagnostic tests. *J Nucl Cardiol* 23(3):511–513. <https://doi.org/10.1007/s12350-015-0175-7>
8. Lin LI (1989) A concordance correlation coefficient to evaluate reproducibility. *Biometrics* 45(1):255–268
9. McBride GB (2020) A Proposal for Strength-of-Agreement Criteria for Lin's Concordance Correlation Coefficient. NIWA Client Report: HAM2005-062; National Institute of Water & Atmospheric Research: Hamilton, New Zealand, May 2005. Available from: <http://www.oalib.com/references/10521225>. Accessed 12 Apr 2020
10. Altman DG (1991) Practical statistics for medical research. Chapman and Hall, London
11. The R Project for Statistical Computing [homepage on the Internet]. Vienna: The R Foundation; c2020 [updated 2020 Apr 24, cited 2020 Apr 24]. R: A Language and Environment for Statistical Computing. <https://cran.rproject.org/doc/manuals/fullrefman.pdf>
12. Bland JM, Altman DG (1999) Measuring agreement in method comparison studies. *Stat Methods Med Res* 8(2):135–160. <https://doi.org/10.1177/096228029900800204>

13. Austin PC, Hux JE (2002) A brief note on overlapping confidence intervals. *J Vasc Surg* 36(1):194–195. <https://doi.org/10.1067/mva.2002.125015>
14. Goldstein H, Healy MJ (1995) The graphical presentation of a collection of means. *J R Stat Soc Ser A* 158(1):175
15. Payton ME, Greenstone MH, Schenker N (2003) Overlapping confidence intervals or standard error intervals: what do they mean in terms of statistical significance? *J Insect Sci* 3:34. <https://doi.org/10.1093/jis/3.1.34>
16. Lopez L, Frommelt PC, Colan SD et al (2020) Pediatric heart network echocardiographic Z scores: Comparison with other published models. *J Am Soc Echocardiogr*. <https://doi.org/10.1016/j.echo.2020.09.019>
17. Lorenzoni R, Elkins N, Quezada M, Silver E, Mahgerefteh J, Choueiter N (2020) Coronary artery Z score formula variability in children with kawasaki disease: pediatric heart network vs. other formulas [abstract 1087-07]. Poster presented at ACC.20/WCC Virtual Conference; 28 Mar 2020
18. Ryu K, Yu JJ, Jun HO, Shin EJ, Heo YH, Baek JS, Kim YH, Ko JK (2017) A comparative study of established z score models for coronary artery diameters in 181 healthy Korean children. *Korean J Pediatr* 60(11):373–378. <https://doi.org/10.3345/kjp.2017.60.11.373>
19. Ronai C, Hamaoka-Okamoto A, Baker AL, de Ferranti SD, Colan SD, Newburger JW, Friedman KG (2016) Coronary artery aneurysm measurement and z-score variability in Kawasaki disease. *J Am Soc Echocardiogr* 29(2):150–157. <https://doi.org/10.1016/j.echo.2015.08.013>
20. Son MBF, Gauvreau K, Tremoulet AH et al (2019) Risk model development and validation for prediction of coronary artery aneurysms in Kawasaki disease in a North American Population. *J Am Heart Assoc* 8(11):e011319. <https://doi.org/10.1161/JAHA.118.011319>
21. Newburger JW, Sleeper LA, McCrindle BW et al (2007) Randomized trial of pulse steroid therapy for primary treatment of Kawasaki disease. *N Engl J Med* 356:663–675. <https://doi.org/10.1056/NEJMoa061235>

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.