Title

Subgroups of children with Kawasaki disease: a data-driven cluster analysis.

Journal

The Lancet Child & Adolescent Health, 7(10)

Authors

Wang, Hao

Shimizu, Chisato

Bainto, Emelia

et al.

Publication Date

2023-10-01

DOI

10.1016/S2352-4642(23)00166-9

Peer reviewed

HHS Public Access

Author manuscript

Lancet Child Adolesc Health. Author manuscript; available in PMC 2023 December 29.

Published in final edited form as:

Lancet Child Adolesc Health. 2023 October ; 7(10): 697–707. doi:10.1016/S2352-4642(23)00166-9.

Subgroups of Kawasaki disease patients: A data-driven cluster analysis

Hao Wang[1,*], **Chisato Shimizu**[1], **Emelia Bainto**[1], **Shea Hamilton**[2], **Heather R. Jackson**[2], **Diego Estrada-Rivadeneyra**[2], **Myrsini Kaforou**[2], **Michael Levin**[2], **Joan M. Pancheri**[1,3], **Kirsten B. Dummer**[1,3], **Adriana H. Tremoulet**[1,3], **Jane C. Burns**[1,3,*]

[1]Kawasaki Disease Research Center, Department of Pediatrics, University of California San Diego, La Jolla, CA, United States

[2]Section of Paediatric Infectious Disease, Department of Infectious Disease, Imperial College London, London, United Kingdom

[3]Rady Children's Hospital-San Diego, San Diego, CA, United States

SUMMARY

Background—Although Kawasaki disease is commonly regarded as a single disease entity, variability in clinical manifestations and disease outcome has been recognised. We aimed to use a data-driven approach to identify clinical subgroups.

Methods—We analysed clinical data from patients with Kawasaki disease diagnosed at Rady Children's Hospital (San Diego, CA, USA) between Jan 1, 2002, and June 30, 2022. Patients were grouped by hierarchical clustering on principal components with k-means parcellation based on 14 variables, including age at onset, ten laboratory test results, day of illness at the first intravenous immunoglobulin infusion, and normalised echocardiographic measures of coronary artery diameters at diagnosis. We also analysed the seasonality and Kawasaki disease incidence from 2002 to 2019 by subgroup. To explore the biological underpinnings of identified subgroups, we did differential abundance analysis on proteomic data of 6481 proteins from 32 patients with Kawasaki disease and 24 healthy children, using linear regression models that controlled for age and sex.

Findings—Among 1016 patients with complete data in the final analysis, four subgroups were identified with distinct clinical features: (1) hepatobiliary involvement with elevated alanine transaminase, gamma-glutamyl transferase, and total bilirubin levels, lowest coronary artery aneurysm but highest intravenous immunoglobulin resistance rates (n=157); (2) highest band neutrophil count and Kawasaki disease shock rate (n=231); (3) cervical lymphadenopathy with high markers of inflammation (erythrocyte sedimentation rate, C-reactive protein, white blood cell,

*Correspondence authors.

Contributors

HW and JCB conceptualised and designed the study, interpreted the results, and wrote the manuscript. HW did the analysis. JCB, AHT, KBD, and JMP enrolled patients and collected data. CS and EB collected and curated the clinical data. SH, HRJ, DE-R, MK, and ML collected and curated the proteomic data. HW and JCB accessed and verified the data. All authors reviewed the results, revised the manuscript, had access to all the data in this study and participated in the decision to submit.

Declaration of interests

We declare no competing interests.

and platelet counts) and lowest age-adjusted haemoglobin *Z* scores (n=315); and (4) young age at onset with highest coronary artery aneurysm but lowest intravenous immunoglobulin resistance rates (n=313). The subgroups had distinct seasonal and incidence trajectories. In addition, the subgroups shared 211 differential abundance proteins while many proteins were unique to a subgroup.

Interpretation—Our data-driven analysis provides insight into the heterogeneity of Kawasaki disease, and supports the existence of distinct subgroups with important implications for clinical management and research design and interpretation.

INTRODUCTION

Kawasaki disease is an acute paediatric vasculitis and the leading cause of acquired cardiovascular disease in children in high-income countries.[1] The disease was first described in 1967 on the basis of clinical signs and symptoms among 50 Japanese children, including an initial presentation of high fever, mucocutaneous inflammation, and cervical lymphadenopathy.[2] Subsequently, an epidemiological survey identified coronary artery aneurysms as the most concerning complication of Kawasaki disease.[3] Although the pathogenesis of Kawasaki disease has not been fully elucidated, it most likely involves a complex interplay between genetic susceptibility and various environmental triggers.[4,5]

Emerging evidence supports a heterogeneous nature of Kawasaki disease, despite being commonly referred to as a single disease entity. Over the past five decades, marked variability in clinical manifestations of Kawasaki disease has been reported. The 2017 American Heart Association (AHA) guidelines on Kawasaki disease listed 21 other clinical findings affecting seven major organ systems that were not described in the original case definition for Kawasaki disease.[6] Furthermore, studies on disease outcome have revealed varied risks of coronary artery aneurysms associated with young age at onset, Asian ethnicity, male sex, laboratory findings of increased inflammation, and early coronary artery dilation.[7,8] In addition, a 2021 multiomic study uncovered transcriptomic and proteomic signatures that categorised patients with Kawasaki disease into three different groups, suggesting that the distinct host responses resulted from different triggers interacting with varied host genetics.[9]

In clinical practice, patients are classified as either complete or incomplete Kawasaki disease, as defined by the number of features at presentation meeting the widely used diagnostic criteria.[6] Accurate description of Kawasaki disease subgroups might contribute to stratified clinical management and prognostication of risk of coronary artery aneurysm. It might also benefit research by disentangling the subgroup-specific pathophysiology that was previously not considered. Therefore, we aimed to do an unsupervised, data-driven cluster analysis to profile the subgroups of patients with Kawasaki disease and explore their biological underpinnings.

METHODS

Study design and participants

For this cluster analysis, 1346 patients diagnosed and treated at Rady Children's Hospital-San Diego with onset of Kawasaki disease between Jan 1, 2002, and June 30, 2022, were identified from the REDCap database at the Kawasaki Disease Research Center, University of California San Diego (UCSD; San Diego, CA, USA). The diagnosis was made by two highly experienced Kawasaki disease specialists (JCB and AHT) on the basis of the AHA definition for either complete or incomplete Kawasaki disease.[6] Clinical findings were prospectively recorded on standardised data collection forms at the time of diagnosis by these two clinicians for all patients. In the first 2 months after disease onset, echocardiography was done at diagnosis, at the 2-week clinic visit, and more frequently if clinically indicated. During the study period, there were no significant changes in data collection, laboratory testing, or treatment, except for the introduction of infliximab in 2014 for intravenous immunoglobulin resistance or abnormal coronary artery dimensions on the first echocardiogram at diagnosis. Only patients with complete data for the selected variables of interest were included in the final cluster analysis. We compared the patients with complete data in the study cohort with those excluded due to missing data; due to the concern of including patients with SARS-CoV-2-associated multisystem inflammatory syndrome in children (MIS-C), we further compared the patients who were diagnosed before and during the pandemic period (2020–22) within the study cohort. Parents and patients signed informed consent or assent as appropriate prospectively at the time of diagnosis for this study, which was approved by the UCSD Institutional Review Board (number 140220).

Comparison of clinical features

Demographic features (age at onset, sex, and race), physical findings associated with Kawasaki disease, relevant laboratory test results (complete blood count with manual differentials, chemistries, erythrocyte sedimentation rate [ESR], and C-reactive protein level [CRP]) at diagnosis, and days of illness at first dose of intravenous immuno- globulin infusion were retrieved. Echocardiographic measures of coronary artery internal diameters both at diagnosis and during follow up were included. Haemoglobin concentrations were normalised for age, and coronary artery internal dimensions were adjusted for body surface area using the Dallaire equation.[10]

The diagnostic criteria of Kawasaki disease shock syndrome were previously described, including sustained presence of any of the following conditions that prompted initiation of volume expansion, vasoactive agents, or transfer to an intensive care setting: systolic hypotension for age, a decrease in systolic blood pressure from baseline of 20% or more, or clinical signs of poor perfusion regardless of measured blood pressure.[11]

The coronary artery dimensions were classified on the basis of the AHA definition using the maximum Z scores for the right and left anterior descending coronary arteries within 60 days from the onset of fever: normal (Z<2), dilation (Z≥2 to <2·5), and aneurysm (Z≥2·5).[6] Intravenous immunoglobulin resistance was defined as persistent or recrudescent fever at least 36 h following completion of the first infusion of intravenous immunoglobulin.[6]

Clinical features were compared by the more conservative non-parametric Kruskal-Wallis (for continuous variables) or χ^2 test (for categorical variables) among all subgroups, and post hoc Dunn's (for continuous variables) or χ^2 test (for categorical variables) between any of the two subgroups. The Benjamini-Hochberg method was used to correct for multiple comparisons. Statistical significance was defined as a p value or false discovery rate less than 0·05. The statistical tests were computed with the R packages FSA version 0.9.4 and chisq.posthoc.test version 0.1.2, and the ggplot2 version 3.4.2 was used for data visualisation. All analyses were done with R version 4.2.2.

Cluster analysis of clinical phenotypes

We did a two-step procedure to select phenotypic variables for cluster analysis: first, to avoid potential bias and preserve sample size, we excluded those data elements with a missing rate above 20%, including aspartate transaminase level, *Z* scores of the left anterior descending and right coronary artery measures before discharge, and at the 2-week clinic visit (appendix pp 2–3); second, we identified highly correlated pairs of variables (Pearson's r>0·8) and excluded one variable from each pair, including day of illness at diagnosis, absolute band and neutrophil counts (appendix pp 2–3). This resulted in the selection of 14 continuous variables for subsequent cluster analysis, including one demographic feature (age at onset), ten laboratory results (pre-treatment, age-adjusted *Z* score of haemoglobin, white blood cell and platelet counts, percentages of neutrophils, bands, and lymphocytes, alanine transaminase, gamma-glutamyl transferase, ESR, and CRP), and the body surface area-adjusted *Z* scores for the left anterior descending artery and right coronary artery at diagnosis. To test our hypothesis of clustered clinical features among patients with Kawasaki disease, we calculated Hopkin's statistic with the R package factoextra version 1.0.7 to quantify the clustering tendency of phenotypic data. A value close to one indicates that the data are highly clustered, random data will tend to result in values around 0·5, and uniformly distributed data will result in values close to zero.[12]

We used hierarchical clustering on principal components for cluster analysis, which involves normalisation of input data, principal component analysis, hierarchical clustering, parcellation of hierarchical tree by the optimal number of clusters, and result consolidation by *k*-means clustering (an algorithm that partitions the samples into *k* clusters based on smallest mean distances to cluster centres).[13] Principal component analysis was done on the 14 selected continuous variables (appendix p 4). All categorical variables and maximum *Z* score of all coronary artery measurements within 60 days from onset of fever were used as supplementary variables that neither participated in dimensionality reduction nor affected the cluster results, but only assisted in interpretation of the clinical characteristics of the identified clusters. The optimal number of clusters was determined by comparing the inertia, which measures the sum of squared distance of samples to their closest cluster centre and quantifies how well a dataset is clustered. The cluster analysis was computed by the FactoMineR package version 2.7 for R.[14]

Epidemiological data and analysis

We studied the seasonality and year-to-year changes in incidence of the identified Kawasaki disease subgroups by plotting their monthly and yearly proportions of newly diagnosed

cases. To study the year-to-year changes, we included the years of 2002–21 with complete data of all 12 months. For the analysis of monthly changes, 2020 and 2021 were not included in view of the reported change in Kawasaki disease incidence during the COVID-19 pandemic.[15]

Proteomic data and analysis

We acquired the pre-treatment proteomic data from 32 patients with Kawasaki disease from our centre and 24 afebrile, healthy children as the controls from the PERFORM study. Briefly, the proteomic abundance was measured from pre-intravenous immunoglobulin treatment, acute phase plasma samples by the SomaScan (SomaLogic, Boulder, CO, USA) aptamer-based platform.16 Quality control steps used scale factors returned from the SomaScan platform to correct for variations in aptamer hybridisation efficiency, inter-assay and intra-assay variability, variability in the starting quantities of proteins, and plate effects.[9]

Kawasaki disease-associated protein differential abundance—which detects changes in protein levels in order to reveal difference in pathophysiology among the Kawasaki disease subgroups—between the healthy control group and each identified Kawasaki disease subgroup, and subgroup-related protein differential abundance comparing all subgroups, were analysed using linear regression models that controlled for effects of age and sex. To reveal the tissue origin of the identified subgroup-related differential abundance proteins among the subgroups, their Ensembl gene symbols were used for tissue enrichment analysis using Enrichr, a web-based tool, and Human U133A/ GNF1H Gene Atlas from BioGPS as reference.[17,18]

Role of the funding source

The funder of the study had no role in study design, data collection, data analysis, data interpretation, or writing of the report.

RESULTS

Of 1346 patients identified from the database, 1016 had complete data for the selected variables of interest and were included in the final analysis. We first evaluated our Kawasaki disease subgroup hypothesis by quantifying the clustering tendency of the phenotypic data. The calculated Hopkin statistic was 0·836, which indicated a highly clustered structure and the possible existence of patient subgroups. Next, the optimal number of clusters was determined as k=4 on the basis of the calculated inertia (figure 1). Furthermore, dimensionality reduction was done by principal component analysis. Principal component one (19·48% of variance) captured the variance due to baseline coronary artery measurement, age at onset, and hepatobiliary biomarker levels, whereas Principal component two (14·69% of variance) largely explained the variance due to inflammation-associated biochemical and haematological changes (figure 1; appendix p 4).

Based on the above results, four patient subgroups were identified following the algorithms of hierarchical clustering on principal components with distinct clinical features (table 1): (1) hepatobiliary involvement with elevated alanine transaminase, gamma-glutamyl transferase, and total bilirubin levels; (2) high band neutrophil count and Kawasaki

disease shock rate; (3) highest rate of cervical lymphadenopathy, elevated markers of inflammation (high ESR, CRP, white blood cell, and platelet counts), and lowest age-adjusted haemoglobin *Z* scores; and (4) young age at onset of Kawasaki disease. To summarise their main features and facilitate further discussion, we named them the liver, band, node, and young subgroups, respectively. The detailed clinical characteristics of the study cohort (n=1016) are shown in table 2.

We observed marked differences in the age at onset of Kawasaki disease among the identified subgroups of patients, with the earliest age at onset in the young subgroup and older age at onset in the liver subgroup. Although no significant over-representation or under- representation of any specific race in any subgroup was found, a χ^2 test indicated unbalanced distribution of racial and ethnic groups (p=0·036; table 2; appendix pp 5–6).

Nearly half of the children in the node subgroup developed significant cervical lymphadenopathy (p<0·0001), a rarer presentation in both the band (post- hoc χ^2 test p=0·050) and young (post-hoc χ^2 test p<0·0001) subgroups. In addition, fewer patients in the young subgroup had documented oral mucosal changes (p=0·0010), consistent with the clinical observation of incomplete and more subtle Kawasaki disease manifestations among younger patients.19 In addition, a significantly higher proportion of children in the band subgroup presented with Kawasaki disease shock syndrome (p<0·0001; table 2; appendix pp 5–6).

There were significant subgroup-related variations in disease outcome and intravenous immunoglobulin resistance (appendix pp 5–6). Risk of coronary artery involvement was highest in the young subgroup, including the largest left anterior descending artery and right coronary artery *Z* scores at diagnosis, and the highest rate of coronary artery aneurysm formation. We found that the distribution of coronary artery aneurysm remained the same among the subgroups of patients before or after the introduction of infliximab treatment for intravenous immunoglobulin resistance or abnormal coronary artery dimensions on the first echocardiogram at diagnosis (appendix p 7). The above risk followed a gradient pattern across the four subgroups, with the lowest risk in the liver subgroup, which roughly correlated with the age difference among the four subgroups. A reverse pattern of intravenous immunoglobulin resistance was also observed, with the highest rate in the liver subgroup (post-hoc χ^2 test p=0·0032) and lowest in the young subgroup (post-hoc χ^2 test p=0·0006). Overall, patients in the young subgroup were more likely to be diagnosed with incomplete Kawasaki disease on the basis of abnormal echocardiographic findings (p=0·014; table 2).

Subgroup-specific features were shown on the laboratory test results at diagnosis (appendix p 8). Prominent liver involvement, manifested as elevated alanine transaminase, gamma-glutamyl transferase, and total bilirubin, was the key feature of the liver subgroup. The per quintile levels of these three hepatobiliary biomarkers were positively related to the odds ratios for intravenous immunoglobulin resistance among all patients in the study cohort (appendix p 9). There was profound inflammation in the node subgroup, including high white blood cell and platelet counts, highest ESR and CRP levels, and low age-adjusted haemoglobin concentration. Patients in the young subgroup presented with low neutrophil

and high lymphocyte percentages, possibly related to their young age-related physiology. Although white blood cell count was lower in the band subgroup, there was a remarkable increase in immature band neutrophils in this subgroup, with both the highest percentage and absolute count among all four subgroups (table 2). We observed no significant difference in sex and age at onset between patients who were excluded from the analysis due to incomplete data and those included in the analysis (appendix p 13). Similarly, when comparing patients diagnosed before and during the COVID-19 pandemic period (2020–22) using all available data, no significant difference was found (appendix pp 14–15).

Next, we studied the seasonality (2002–19) and year-to-year changes (2002–21) in new case proportions of the identified Kawasaki disease subgroups, and observed different trajectories. Although the proportion of patients in the liver subgroup was largely stable throughout the year, a summer trough was observed for patients in the young subgroup, compared with those of the band and node subgroups that peaked in August and July, respectively. There were additional winter peaks of new cases in patients in the node and young subgroups that were absent in the other two subgroups (figure 2A).

We also showed year-to-year differences over the past two decades, including an overall rise in proportions of patients in the liver subgroup and concomitant decrease in patients in the band subgroup. The changes were more obvious during the years affected by the COVID-19 pandemic in 2020–21 (figure 2B). In addition, there were notable changes in subgroup-related seasonality when comparing the two decades (2002–10 and 2011–19) with the winter peak of patients in the young subgroup shifting from December to March, and a new increase in the proportion of patients in the liver subgroup in October (appendix p 10).

Finally, we explored the biological variations that might underlie the identified Kawasaki disease subgroups with proteomic differential abundance analysis. There was no significant difference in demographic features, clinical presentation, or laboratory test results between the proteomic study cohort (n=32) and the rest of the cluster analysis cohort (n=984), except for a lower frequency of documented cervical lymphadenopathy (p=0·030) and lower white blood cell counts (p=0·0043) among the 32 patients (appendix pp 16–17). First, we compared each subgroup of 32 patients with Kawasaki disease (liver n=5, band n=9, node n=9, young n=9) with a control group of 24 healthy children, and identified 211 Kawasaki disease- associated differential abundance proteins that were shared among the four subgroups after excluding age and sex effects, including signatures of transforming growth factor β signalling. There were also 135, 264, 613 and 46 differential abundance proteins unique to the liver, band, node, and young subgroups, respectively (figure 3A), implying subgroup-related variations from a common Kawasaki disease pathophysiology.

Further, a comparison among the four subgroups identified 40 subgroup-related differential abundance proteins (appendix pp 18–19). Most were highest in the liver subgroup (figures 3B–C; appendix p 11), suggesting underlying hepatobiliary injury. Several other differential abundance proteins might suggest unique underlying pathophysiology of their most abundant subgroups, such as IL-18 receptor 1 (IL-18 Rα) for the liver subgroup, soluble urokinase plasminogen activator receptor (suPAR) for the band subgroup, and single immunoglobulin IL-1-related receptor (SIGIRR) for the young subgroup (figure 3B).

DISCUSSION

We evaluated the phenotypic heterogeneity of Kawasaki disease with an objective, data-driven approach. The calculated clustering tendency and the distinct clinical features of the identified four subgroups of patients support the existence of Kawasaki disease subgroups. We further observed different trajectories of both monthly and yearly new case proportions among the identified subgroups, suggesting different environmental triggers. Finally, we examined the underlying pathophysiology and showed proteome-level evidence supporting the Kawasaki disease subgroups. Together, these results depict a model that involves intricate interactions between different triggers and genetic susceptibility, leading to activation of a shared disease pathway (eg, transforming growth factor β signalling and downstream cascades[20]) and unique pathways that vary among patients, resulting in the distinct clinical subgroups (figure 4).

The young subgroup with the earliest onset of Kawasaki disease and the liver subgroup with hepatobiliary involvement most clearly represent two distinct subgroups. Considering the gradient of age across these subgroups, age-related differences in immune response and exposure to environmental triggers might contribute to the observed variations. Discernible clinical features of patients with a young age at onset of Kawasaki disease, especially younger than 6 months, have been previously reported, including incomplete diagnostic criteria and a high risk of coronary artery aneurysm formation even with timely treatment,[7,19] all of which are consistent with the phenotypic characteristics of the young subgroup identified in our study.

Clinical manifestations of liver involvement in Kawasaki disease range from asymptomatic elevation of hepatobiliary enzymes to severe cholestatic hepatitis or gallbladder hydrops.[21] Portal triad inflammation has been described in liver biopsies and from autopsy studies.[22–24] The present study also showed a large number of differential abundance proteins of hepatic origin, most likely leaked from injured hepatocytes. Moreover, we noted the highest intravenous immuno- globulin-resistance rate in this subgroup, and an overall dose–response relationship between the risk of intravenous immunoglobulin resistance and levels of alanine transaminase, gamma-glutamyl transferase, and total bilirubin among all patients. These findings are in line with previous observations that elevated hepatobiliary enzymes and specifically gamma-glutamyl transferase are associated with resistance to intravenous immunoglobulin.[25,26] This is also consistent with the previous discovery in a Japanese cohort that total bilirubin was an independent predictor of intravenous immunoglobulin resistance,[27] as hyperbilirubinaemia is part of the clinical manifestations of hepatobiliary injury. In addition, we observed a unique incidence trajectory of this Kawasaki disease subgroup, particularly in 2020–21 during the COVID-19 pandemic when there was a sharp decrease in both absolute number and proportion of new cases in this subgroup. The findings could imply a unique environmental trigger of hepatobiliary injury in Kawasaki disease that was reduced by behaviours, such as masking and physical distancing, or environmental changes, such as reduced nitric oxide pollutants from the transportation sector.[15]

Three differential abundance proteins were highly abundant in the liver (IL-18 Rα), young (SIGIRR), and band (suPAR) subgroups and could hint at possible subgroup-specific

biological pathways. IL-18 Rα and SIGIRR are coreceptors of the anti-inflammatory IL-37 and components of the IL-37–IL-18Rα–SIGIRR tripartite complex that blocks binding of IL-18,[28] a key player involved in endothelial cell pyroptosis.[29] As both IL-18 Rα and SIGIRR are required to support the anti- inflammatory activity of IL-37,[28] an imbalance in expression of these two proteins could lead to persistent inflammation with failure to efficiently bind to IL-37.

suPAR is the product from cleavage of the membrane-bound uPAR, and shed from several immunologically active cells, including neutrophils, monocytes, macrophages, and activated T cells.[30] We observed the lowest suPAR levels among healthy children and patients with the youngest age at Kawasaki disease onset and highest coronary artery aneurysm risk, and the highest levels in the band subgroup with the highest Kawasaki disease shock rate. The significant difference could imply varied pathophysiology in these Kawasaki disease subgroups.

We recognise several limitations to our study. As an early attempt based on the data from a single centre, the results of the present study might not represent the optimal classification of Kawasaki disease subgroups. Due to limited study power and coverage of patient subpopulations, such as African American children, we were unable to further profile the sub-structures of each subgroup to exhaustively unveil all Kawasaki disease subgroups or show more detail about the subgroup- specific pathophysiology. Future efforts are required to validate our findings using data from patients with Kawasaki disease from other geographic locations to improve the resolution of clustering and delineate the triggers that underlie the Kawasaki disease subgroups.

The existence of subgroups should be considered in research initiatives exploring different aspects of Kawasaki disease. Failure to account for the subgroup- related variance will inevitably distort the effect sizes of research targets of interest, such as molecular and genetic variants and the search for etiological agents. Biases and confounders might also be introduced by using patient cohorts of randomly admixed subgroups, thus impeding the replication and generalisability of conclusions. Therefore, we suggest a subgroup-aware approach in further Kawasaki disease research to minimise subgroup-incurred variance.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

This study was funded by the US National Institutes of Health PreVAIL grant (R61HD105590) and the Irving and Francine Suknow Foundation. We thank all the patients who participated in this trial and their families.

Funding

US National Institutes of Health and the Irving and Francine Suknow Foundation.

Data sharing

Deidentified participant data that underlie the results reported in this Article can be shared upon specific requests. Requests will be honoured from researchers who provide a methodologically sound proposal and who execute a Data Use Agreement with the University of California San Diego. Requests should be directed by email to the corresponding author.

REFERENCES

1. Newburger JW, Takahashi M, Burns JC. Kawasaki disease. J Am Coll Cardiol 2016; 67: 1738–49. [PubMed: 27056781]
2. Kawasaki T, Kosaki F, Okawa S, Shigematsu I, Yanagawa H. A new infantile acute febrile mucocutaneous lymph node syndrome (MLNS) prevailing in Japan. Pediatrics 1974; 54: 271–76. [PubMed: 4153258]
3. Kato H, Koike S, Yamamoto M, Ito Y, Yano E. Coronary aneurysms in infants and young children with acute febrile mucocutaneous lymph node syndrome. J Pediatr 1975; 86: 892–98. [PubMed: 236368]
4. Onouchi Y The genetics of Kawasaki disease. Int J Rheum Dis 2018; 21: 26–30. [PubMed: 29152908]
5. Manlhiot C, Mueller B, O'Shea S, et al. Environmental epidemiology of Kawasaki disease: linking disease etiology, pathogenesis and global distribution. PLoS One 2018; 13: e0191087. [PubMed: 29415012]
6. McCrindle BW, Rowley AH, Newburger JW, et al. Diagnosis, treatment, and long-term management of Kawasaki disease: a scientific statement for health professionals from the American Heart Association. Circulation 2017; 135: e927–99. [PubMed: 28356445]
7. Son MBF, Gauvreau K, Tremoulet AH, et al. Risk model development and validation for prediction of coronary artery aneurysms in Kawasaki disease in a North American population. J Am Heart Assoc 2019; 8: e011319. [PubMed: 31130036]
8. Dominguez SR, Anderson MS, El-Adawy M, Glodé MP. Preventing coronary artery abnormalities: a need for earlier diagnosis and treatment of Kawasaki disease. Pediatr Infect Dis J 2012; 31: 1217–20. [PubMed: 22760536]
9. Jackson H, Menikou S, Hamilton S, et al. Kawasaki disease patient stratification and pathway analysis based on host transcriptomic and proteomic profiles. Int J Mol Sci 2021; 22: 5655. [PubMed: 34073389]
10. Dallaire F, Dahdah N. New equations and a critical appraisal of coronary artery Z scores in healthy children. J Am Soc Echocardiogr 2011; 24: 60–74. [PubMed: 21074965]
11. Kanegaye JT, Wilder MS, Molkara D, et al. Recognition of a Kawasaki disease shock syndrome. Pediatrics 2009; 123: e783–89. [PubMed: 19403470]
12. Hopkins B, Skellam JG. A new method for determining the type of distribution of plant individuals. Ann Bot (Lond) 1954; 18: 213–27.
13. Argüelles M, Benavides C, Fernández I. A new approach to the identification of regional clusters: hierarchical clustering on principal components. Appl Econ 2014; 46: 2511–19.
14. Lê S, Josse J, Husson F. FactoMineR: An R package for multivariate analysis. J Stat Softw 2008; 25: 1–18.
15. Burney JA, Roberts SC, DeHaan LL, et al. Epidemiological and clinical features of Kawasaki disease during the COVID-19 pandemic in the United States. JAMA Netw Open 2022; 5: e2217436. [PubMed: 35713905]
16. Gold L, Ayers D, Bertino J, et al. Aptamer-based multiplexed proteomic technology for biomarker discovery. PLoS One 2010; 5: e15004. [PubMed: 21165148]
17. Kuleshov MV, Jones MR, Rouillard AD, et al. Enrichr: a comprehensive gene set enrichment analysis web server 2016 update. Nucleic Acids Res 2016; 44: W90–97. [PubMed: 27141961]

18. Su AI, Wiltshire T, Batalov S, et al. A gene atlas of the mouse and human protein-encoding transcriptomes. Proc Natl Acad Sci USA 2004; 101: 6062–67. [PubMed: 15075390]
19. Salgado AP, Ashouri N, Berry EK, et al. High risk of coronary artery aneurysms in infants younger than 6 months of age with Kawasaki disease. J Pediatr 2017; 185: 112–116.e1. [PubMed: 28408126]
20. Shimizu C, Jain S, Davila S, et al. Transforming growth factor-beta signaling pathway in patients with Kawasaki disease. Circ Cardiovasc Genet 2011; 4: 16–25. [PubMed: 21127203]
21. Eladawy M, Dominguez SR, Anderson MS, Glodé MP. Abnormal liver panel in acute kawasaki disease. Pediatr Infect Dis J 2011; 30: 141–44. [PubMed: 20861758]
22. Ohshio G, Furukawa F, Fujiwara H, Hamashima Y. Hepatomegaly and splenomegaly in Kawasaki disease. Pediatr Pathol 1985; 4: 257–64. [PubMed: 3835550]
23. Bader-Meunier B, Hadchouel M, Fabre M, Arnoud MD, Dommergues JP. Intrahepatic bile duct damage in children with Kawasaki disease. J Pediatr 1992; 120: 750–52. [PubMed: 1578309]
24. Tanaka T, Koike M, Minami Y. Pathology of liver injury in Kawasaki disease. Shoni-Naika 1984; 16: 2393–97 (in Japanese).
25. Wang Y, Li Z, Hu G, et al. Unique molecular patterns uncovered in Kawasaki disease patients with elevated serum gamma glutamyl transferase levels: implications for intravenous immunoglobulin responsiveness. PLoS One 2016; 11: e0167434. [PubMed: 28002448]
26. Liu L, Yin W, Wang R, Sun D, He X, Ding Y. The prognostic role of abnormal liver function in IVIG unresponsiveness in Kawasaki disease: a meta-analysis. Inflamm Res 2016; 65: 161–68. [PubMed: 26645912]
27. Miyata K, Miura M, Kaneko T, et al. Risk factors of coronary artery abnormalities and resistance to intravenous immunoglobulin plus corticosteroid therapy in severe Kawasaki disease: an analysis of post RAISE. Circ Cardiovasc Qual Outcomes 2021; 14: e007191. [PubMed: 33541111]
28. Molgora M, Barajon I, Mantovani A, Garlanda C. Regulatory role of IL-1R8 in immunity and disease. Front Immunol 2016; 7: 149. [PubMed: 27148268]
29. Jia C, Zhang J, Chen H, et al. Endothelial cell pyroptosis plays an important role in Kawasaki disease via HMGB1/RAGE/ cathespin B signaling pathway and NLRP3 inflammasome activation. Cell Death Dis 2019; 10: 778. [PubMed: 31611559]
30. Thunø M, Macho B, Eugen-Olsen J. suPAR: the molecular crystal ball. Dis Markers 2009; 27: 157–72. [PubMed: 19893210]

RESEARCH IN CONTEXT

Evidence before this study

Although Kawasaki disease (KD) is commonly referred to as a single disease entity, its clinical manifestations and outcome have marked variability. The 2017 American Heart Association (AHA) guidelines on Kawasaki disease listed 21 other clinical findings affecting seven major organ systems that were not described in the original case definition of Kawasaki disease, and risk of coronary artery aneurysms varies among different patient populations. To uncover possible Kawasaki disease subgroups and explore the underlying biological mechanisms, we did a comprehensive review using PubMed to identify the existing evidence published up to Feb 14, 2023. The search terms used included “Kawasaki disease” or “KD”, and “subgroup” or “subtype” or “sub-phenotype” or “cluster”. Notably, only few studies were available, including two studies that showed temporal and geographic clusters of patients with Kawasaki disease, with distinct clinical features, suggesting different triggers, and another multiomic study that showed transcriptomic and proteomic signatures, categorising patients with Kawasaki disease into three different groups on the basis of distinct host responses. Nonetheless, a few review papers mentioned the heterogeneous aetiology of KD and need for patient stratification.

Added value of this study

To our knowledge, this is the first cluster analysis to explore clinical Kawasaki disease subgroups using an objective, data-driven approach. The study identified four patient clusters with distinct clinical features, treatment response, and disease outcomes. The subgroup with hepatobiliary involvement and another with the youngest age at onset of Kawasaki disease had opposite risks of intravenous immunoglobulin treatment resistance and coronary artery aneurysms. The existence of Kawasaki disease subgroups was supported by proteome-level and epidemiological evidence. We propose a Kawasaki disease model that involves an intricate interplay between genetic susceptibility and environmental triggers, leading to a common Kawasaki disease pathophysiology that varies in different patients and results in distinct clinical subgroups.

Implications of all the available evidence

The existence of Kawasaki disease subgroups should be considered by both clinicians and researchers. For older children presenting with elevated hepatobiliary biomarkers and consistent physical findings, a diagnosis of Kawasaki disease should be considered even if the inflammatory marker levels are not markedly elevated. This subgroup is more likely to be resistant to intravenous immunoglobulin treatment, but their risk of coronary artery aneurysm is low. For young children at high risk of coronary artery aneurysm, prompt treatment with intravenous immunoglobulin should remain a priority. In addition, we propose a subgroup-aware approach in Kawasaki disease-related research design and data interpretation to account for subgroup-related variance.

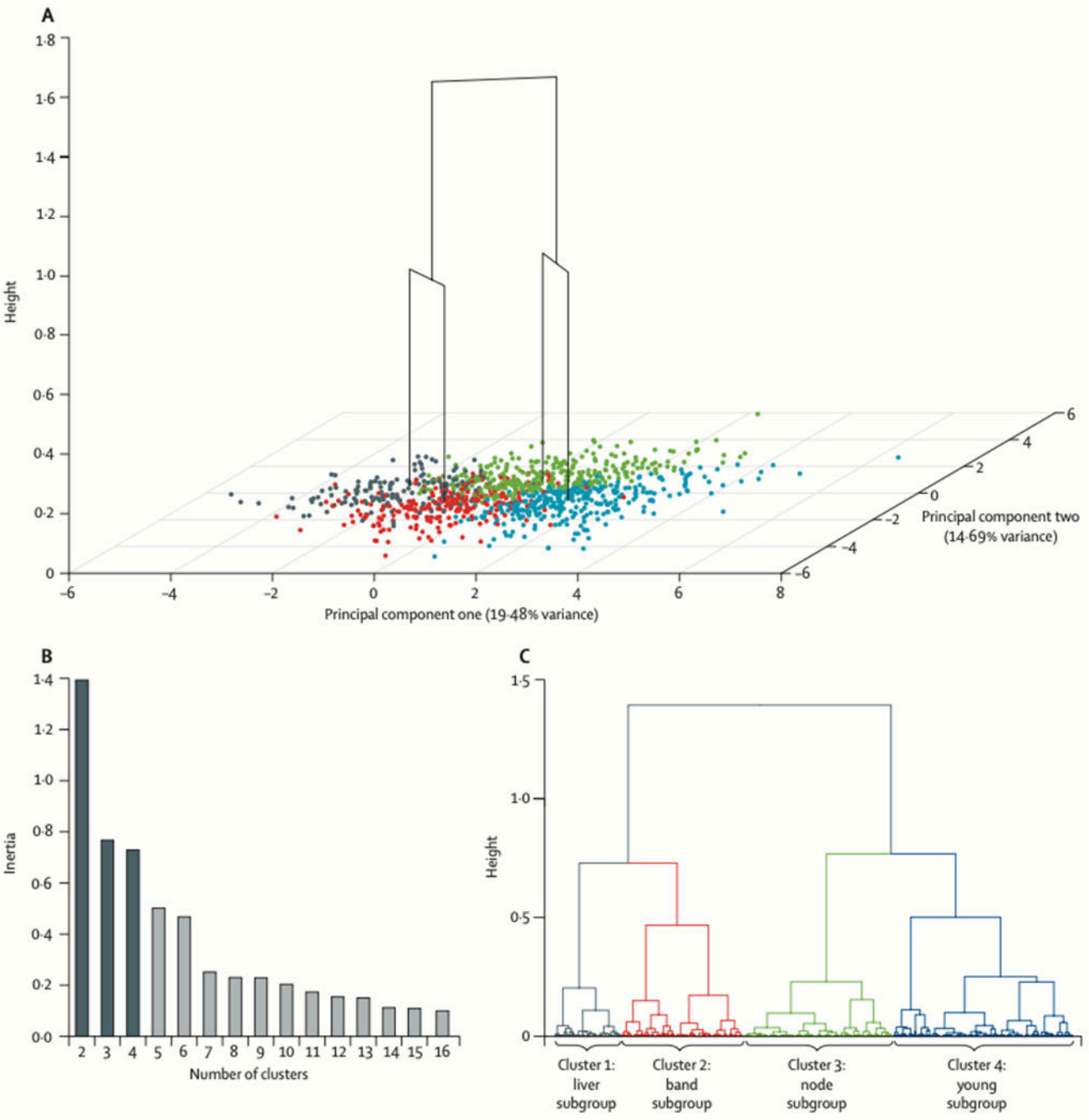

Figure 1. Hierarchical clustering on principal components

(A) Three-dimensional map of hierarchical clustering on principal components. (B) The optimal number of clusters determined by the calculated inertia. (C) Dendrogram of hierarchical map and identified clusters.

Figure 2. Distribution of the Kawasaki disease subgroups

(A) Monthly cumulative numbers (bars) and proportions (lines) of new cases in each phenotypic subgroup in 2002–19. Error bars denote 95% CIs. (B) Yearly numbers (bars) and proportions (lines) of new cases in each phenotypic subgroup in 2002–21. Highlighted are the years affected by the COVID-19 pandemic.

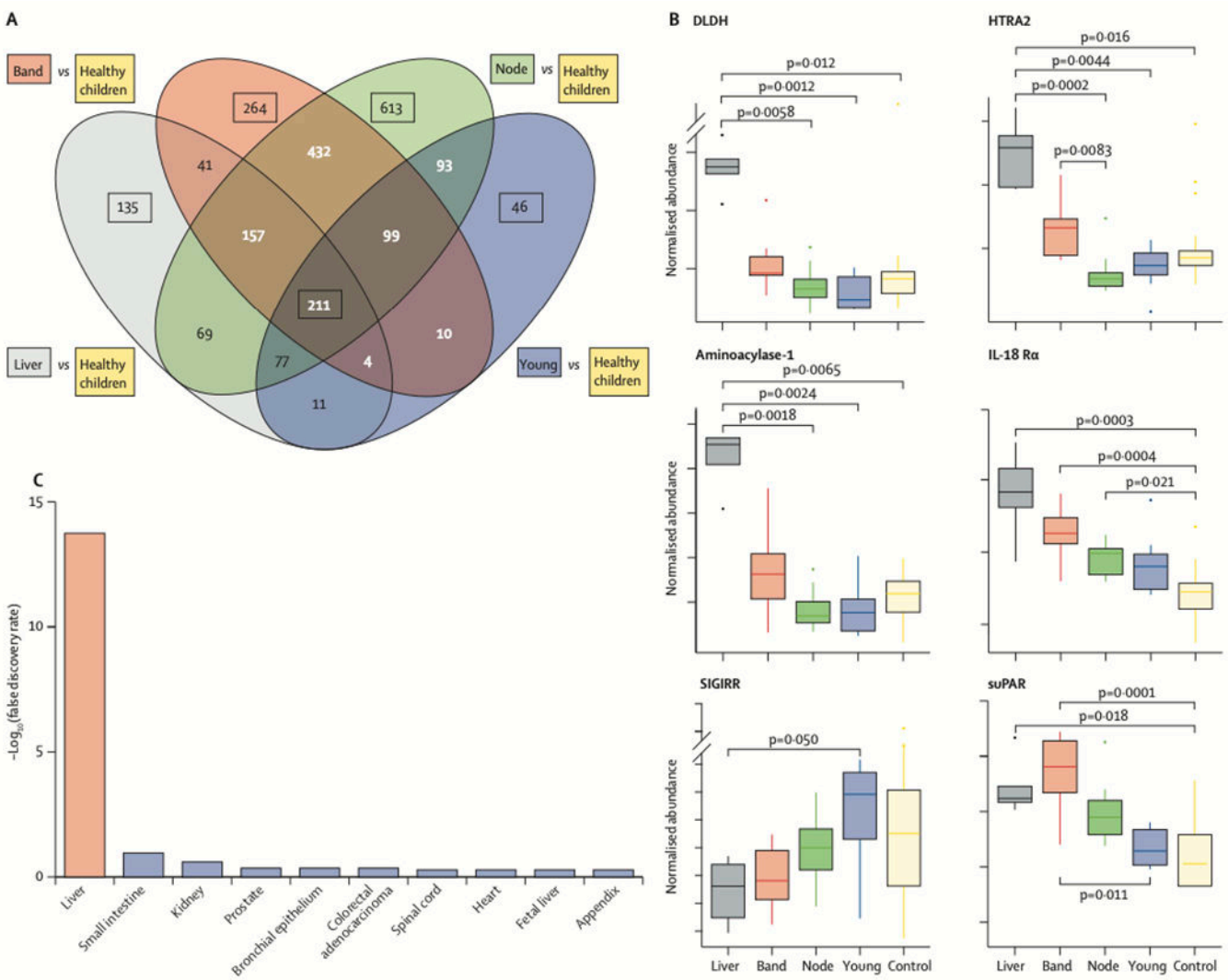

Figure 3. Proteomic analysis

(A) Venn diagram showing the numbers of Kawasaki disease-associated differentially abundant proteins (false discovery rate <0·05) that were shared among the subgroups or unique to a specific subgroup of patients with Kawasaki disease, comparing Kawasaki disease and the control group of 24 healthy children. (B) Selected subgroup-related differential abundance proteins from comparison among the four subgroups. The normalised abundance in the control group (yellow) was also plotted as reference. (C) Tissue enrichment analysis using gene symbols of the 40 subgroup-related differential abundance proteins based on the Human Gene Atlas. Labelled p values were from post hoc Dunn's test for pairwise comparisons. p values were adjusted for multiple comparisons, with the Benjamini-Hochberg method. DLDH=dihydrolipoamide dehydrogenase. HTRA2=serine protease HTRA2. IL-18 Rα=IL-18 receptor 1. suPAR=soluble urokinase plasminogen activator receptor. SIGIRR=single immunoglobulin IL-1-related receptor.

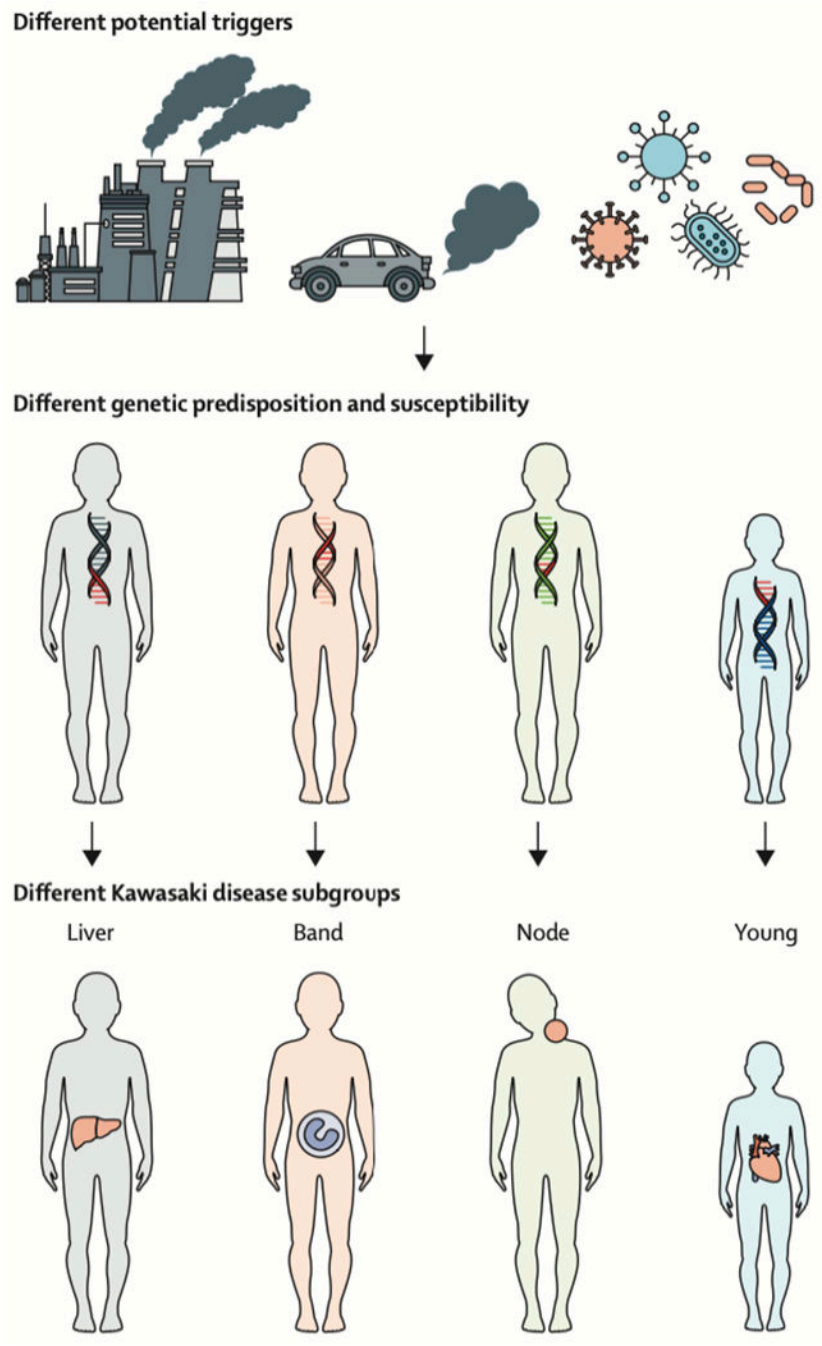

Figure 4. Proposed Kawasaki disease model

Interactions between different environmental triggers and genetic susceptibility lead to activation of a common disease pathway that varies among patients and results in discrete clinical subgroups.

Table 1.

Summary of cluster-specific clinical characteristics.

	Cluster 1 "Liver" (N = 157, 15·5%)	Cluster 2 "Band" (N = 231, 22·7%)	Cluster 3 "Node" (N = 315, 31·0%)	Cluster 4 "Young" (N = 313, 30·8%)
Main feature	**Liver involvement**	**high Band neutrophils**	**cervical lymph Node**	**Young age at onset**
Age at onset	Oldest	Intermediate	Intermediate	Youngest
Incidence of cervical lymphadenopathy	Intermediate	Lower	Highest	Lowest
Laboratory results	↑ ALT, GGT, total bilirubin ↑ Z-Hgb	↑ % bands ↓ WBC, Plts	↑ WBC, Plts ↓ Z-Hgb ↑ ESR, CRP	↑ Plts, % lymphocyte ↓ % neutrophils, % bands ↓ CRP
CAA risk	Lowest	Intermediate	Intermediate	Highest
IVIG-R rate	Highest	Intermediate	Intermediate	Lowest
KD shock rate	Lowest	Highest	Intermediate	Lower

Abbreviations: LN – lymphadenopathy; **CBC** – complete blood count; **Z-Hgb** – age-adjusted Z-score of hemoglobin concentration; **WBC** – white blood cell count; **Plts** – platelet count; **ALT** – alanine transaminase; **GGT** – gamma-glutamyl transferase; **ESR** – erythrocyte sedimentation rate; **CRP** – C-reactive protein; **Z-CA** – Body surface are-adjusted Z-score of coronary artery internal dimension by echocardiography; **CAA** – coronary artery aneurysm; **IVIG-R** – intravenous immunoglobulin resistance.

Table 2.

Demographic features, clinical presentation and clinical laboratory test results of the total cohort and each subgroup.

Feature		Total (N=1,016)	Liver (1) (N=157)	Band (2) (N=231)	Node (3) (N=315)	Young (4) (N=313)	*p*-value
Demographic features							
Sex (N, %)	Male	622 (61·2)	94 (59·9)	141 (61·0)	184 (58·4)	203 (64·9)	0·41
Race/ethnicity * (N, %)	Asian	170 (16·7)	13 (8·2)	41 (17·7)	55 (17·5)	61 (19·5)	0·036
	African American	36 (3·5)	4 (2·5)	7 (3·0)	10 (3·2)	15 (4·8)	
	European Descendent	228 (22·4)	47 (29·9)	61 (26·4)	69 (21·9)	51 (16·3)	
	Hispanic	364 (35·8)	60 (38·2)	75 (32·5)	115 (36·5)	114 (36·4)	
	Mixed	191 (18·8)	29 (18·5)	39 (16·9)	59 (18·7)	64 (20·4)	
	Native American	4 (0·4)	0	3 (1·3)	1 (0·3)	0	
	Pacific Islander	4 (0·4)	0	2 (0·9)	0	2 (0·6)	
	Other	17 (1·7)	3 (1·9)	3 (1·3)	5 (1·6)	6 (1·9)	
Age at onset (median, IQR)		2·7 (1·3-4·5)	4·5 (2·7-7·7)	3·2 (1·7-5·1)	3·3 (2·2-5·1)	1·2 (0·6-2·0)	< 0·0001
Clinical presentation							
Physical findings (N, %)	Rash	928 (91·3)	152 (96·8)	226 (97·8)	269 (85·4)	281 (89·8)	< 0·0001
	Conjunctival injection	938 (92·3)	142 (90·4)	217 (93·4)	291 (92·4)	288 (92·0)	0·73
	Oral mucosal involvement	945 (93·3)	147 (93·6)	222 (96·1)	296 (94·0)	280 (89·5)	0·010
	Cervical lymphadenopathy	337 (33·4)	56 (35·7)	60 (26·0)	149 (47·3)	72 (23·0)	< 0·0001
KD diagnostic criteria based on AHA guidelines (N, %)	Complete	796 (78·3)	130 (82·8)	192 (83·1)	237 (75·2)	237 (75·7)	0·0026
	Incomplete by echo	50 (4·9)	2 (1·3)	8 (3·5)	14 (4·4)	26 (8·3)	
	Incomplete by labs	146 (14·4)	18 (11·5)	24 (10·4)	59 (18·7)	45 (14·4)	
	Incomplete not by AHA criteria	19 (1·9)	5 (3·2)	6 (2·6)	4 (1·3)	4 (1·3)	
Illness day at diagnosis (median, IQR)		5 (4-7)	5 (4-6)	5 (4-5)	6 (5-7)	6 (5-7)	< 0·0001
Coronary artery outcome (N, %)	Normal (Z ** < 2)	632 (62·2)	127 (80·9)	159 (68·8)	197 (62·5)	149 (47·6)	< 0·0001
	Dilation (2 ≤ Z < 2·5)	137 (13·5)	13 (8·3)	29 (12·6)	44 (14·0)	51 (16·3)	
	Aneurysm (Z ≥ 2·5)	247 (24·3)	17 (10·8)	43 (18·6)	74 (23·5)	113 (36·1)	
KD shock syndrome (N, %) (Total N=886)		29 (3·3)	0	19 (9·5)	8 (2·9)	2 (0·7)	< 0·0001
Illness day at 1st IVIG infusion (median, IQR)		6 (5-7)	5 (4-7)	5 (4-6)	6 (5-7)	6 (5-8)	< 0·0001
IVIG resistance (N, %) (Total N=748)		130 (17·4)	36 (28·1)	41 (23·0)	34 (15·0)	19 (8·8)	< 0·0001
Laboratory Results							

Feature	Total (N=1,016)	Liver (1) (N=157)	Band (2) (N=231)	Node (3) (N=315)	Young (4) (N=313)	*p*-value
WBC (10^3/μl)	13·2 (10·3-17·0)	12·3 (9·6-15·5)	9·7 (7·6-12·4)	15·3 (12·2-18·8)	14·1 (11·4-17·7)	< 0·0001
Hemoglobin (Z-score *)**	−1·4 (−2·2,−0·4)	−0·3 (−1·2,0·3)	−1·0 (−1·9,0)	−1·8 (−2·6,−1·0)	−1·6 (−2·4,−0·7)	< 0·0001
Platelets (10^3/μl)	346 (273-434)	304 (259-394)	259 (192-329)	362 (308-444)	412 (339-493)	< 0·0001
Neutrophils (%)	56 (44-67)	66 (55-73)	46 (35-56)	67 (61-74)	46 (36-53)	< 0·0001
Bands (%)	8 (3-17)	9 (3-16)	24 (11-35)	7 (2-13)	5 (1-11)	< 0·0001
Lymphocytes (%)	20 (12-31)	12 (8-18)	18 (12-26)	16 (10-20)	35 (28-44)	< 0·0001
ESR (mm/hr)	59 (38-75)	47 (32-62)	42 (29-60)	72 (56-100)	63 (44-75)	< 0·0001
CRP (mg/dL)	7·0 (4·4-16·4)	5·6 (3·7-8·1)	8·3 (4·9-17·0)	15·0 (6·2-22·0)	5·2 (3·2-7·7)	< 0·0001
ALT (U/L)	46 (25-117)	234 (136-383)	47 (29-106)	35 (22-74)	30 (20-55)	< 0·0001
GGT (U/L)	47 (19-129)	203 (153-267)	51 (18-119)	33 (19-80)	28 (15-63)	< 0·0001
Total bilirubin** (mg/dl)** (Total N = 364)	0·6 (0·4-0·9)	1·2 (0·7-2·6)	0·6 (0·4-1·0)	0·5 (0·4-0·8)	0·5 (0·3-0·6)	< 0·0001
Baseline Z-LAD	1·4 (0·8-2·1)	1·1 (0·4-1·6)	1·2 (0·6-1·9)	1·3 (0·8-2·0)	1·8 (1·1-2·5)	< 0·0001
Baseline Z-RCA	0·7 (0·1-1·5)	0·3 (−0·4-0·9)	0·6 (0·0-1·2)	0·8 (0·1-1·5)	1·1 (0·5-1·9)	< 0·0001
Zmax	1·7 (1·2-2·5)	1·4 (0·9-1·8)	1·6 (1·1-2·2)	1·7 (1·1-2·4)	2·0 (1·6-3·0)	< 0·0001

* Race/ethnicity was abstracted from medical records based on parent report. **Z**: body surface area-adjusted Z-scores of coronary artery internal diameter measures.

** Age-adjusted Z-scores of hemoglobin concentration.

*** Not in PCA due to the high rate of missing data (64·1%).

Abbreviations: **IQR** – interquartile rangae; **WBC** – white blood cells; **ESR** – erythrocyte sedimentation rate; **CRP** – C-reactive protein; **ALT** – alanine transaminase; **GGT** – gamma-glutamyl transferase; **Z-LAD** – Z-score of left anterior descending artery; **Z-RCA** – Z-score of right coronary artery; **Zmax** – maximum Z-score of all coronary artery measurements within 60 days from onset of fever.

All percentages refer to percentages in a subgroup or the total cohort. *P*-values were from Kruskal-Wallis or chi-squared tests for all-subgroup comparisons. Laboratory results are presented as median and IQR.