Traumatic Head Injury and the Diagnosis of Abuse: A Cluster Analysis

Stephen C. Boos, MD,[a] Ming Wang, PhD,[b] Wouter A. Karst, MD,[c] Kent P. Hymel, MD,[d]
Pediatric Brain Injury Research Network (PediBIRN) Investigators

abstract

OBJECTIVES: Data guiding abusive head trauma (AHT) diagnosis rest on case-control studies that have been criticized for circularity. We wished to sort children with neurologic injury using mathematical algorithms, without reference to physicians' diagnoses or predetermined diagnostic criteria, and to compare the results to existing AHT data, physicians' diagnoses, and a proposed triad of findings.

METHODS: Unsupervised cluster analysis of an existing data set regarding 500 young patients with acute head injury hospitalized for intensive care. Three cluster algorithms were used to sort (partition) patients into subpopulations (clusters) on the basis of 32 reliable ($\kappa > 0.6$) clinical and radiologic variables. *P* values and odds ratios (ORs) identified variables most predictive of partitioning.

RESULTS: The full cohort partitioned into 2 clusters. Variables substantially ($P < .001$ and OR $>$ 10 in all 3 cluster algorithms) more prevalent in cluster 1 were imaging indications of brain hypoxemia, ischemia, and/or swelling; acute encephalopathy, particularly when lasting >24 hours; respiratory compromise; subdural hemorrhage or fluid collection; and ophthalmologist-confirmed retinoschisis. Variables substantially ($P < .001$ and OR < 0.10 in any cluster algorithm) more prevalent in cluster 2 were linear parietal skull fracture and epidural hematoma. Postpartitioning analysis revealed that cluster 1 had a high prevalence of physician-diagnosed abuse.

CONCLUSIONS: Three cluster algorithms partitioned the population into 2 clusters without reference to predetermined diagnostic criteria or clinical opinion about the nature of AHT. Clinical difference between clusters replicated differences previously described in comparisons of AHT with non-AHT. Algorithmic partition was predictive of physician diagnosis and of the triad of findings heavily discussed in AHT literature.

Full article can be found online at www.pediatrics.org/cgi/doi/10.1542/peds.2021-051742

[a]*Department of Pediatrics, University of Massachusetts Chan Medical School—Baystate, Baystate Medical Center, Springfield, Massachusetts;* [b]*Department of Public Health Sciences, College of Medicine, Pennsylvania State University, Hershey, Pennsylvania;* [c]*Department of Forensic Medicine, Netherlands Forensic Institute, The Hague, Netherlands; and* [d]*Department of Pediatrics, College of Medicine, Pennsylvania State University and Penn State Health Children's Hospital, Hershey, Pennsylvania*

Dr Boos conceptualized and designed the study, helped capture the data used in the analysis, drafted the manuscript, and revised the manuscript; Dr Wang helped design the study, completed the cluster analysis, drafted portions of the manuscript, and helped revise the manuscript; Dr Karst helped interpret the findings and helped revise the manuscript; Dr Hymel helped conceptualize and design the study, helped capture data used in the analysis, completed preliminary statistical analyses, drafted portions of the manuscript, and helped revise the manuscript; and all authors approved the final manuscript as submitted and agree to be accountable for all aspects of the work.

DOI: https://doi.org/10.1542/peds.2021-051742

Accepted for publication Aug 27, 2021

WHAT'S KNOWN ON THIS SUBJECT: Existing literature on abusive head trauma (AHT) relies on presumptions about abuse diagnosis to separate cases from controls to identify patterns of clinical findings associated with abuse. This literature has been criticized for circular reasoning.

WHAT THIS STUDY ADDS: Partition by 3 cluster algorithms revealed subpopulations among children with head injury without reference to predetermined diagnostic criteria or clinical opinion about the nature of AHT. Clinical difference between clusters replicated differences previously described in comparisons of AHT with non-AHT.

To cite: Boos SC, Wang M, Karst WA, et al. Traumatic Head Injury and the Diagnosis of Abuse: A Cluster Analysis. *Pediatrics*. 2022;149(1):e2021051742

Neurologic injury resulting from inflicted trauma is a substantial cause of morbidity and mortality in infants and young children.[1–13] Clinicians caring for children with neurologic injury have the additional task of recognizing children whose injuries are the result of abusive head trauma (AHT). The scientific foundation for identifying AHT largely rests on a series of case-control studies that compare abused with nonabused children and on meta-analyses of those studies.[14–30] Some authors have suggested that much of this literature is invalid because of circularity. They find that potential indicators of abuse being studied are used in assigning children as cases or controls.[31] Additionally, authors have opined that child abuse pediatricians routinely diagnose AHT solely on the basis of the presence of 3 findings: subdural hemorrhage (SDH), encephalopathy, and retinal hemorrhage (RH), together known as the triad. They assert that the triad lacks sufficient research validation to support this diagnosis.[31–33] These critiques question the possibility of identifying a subpopulation of children with AHT within the larger population of children with neurologic injury, both in research and in clinical practice.

Cluster analysis is a family of exploratory tools used to explore data sets to see how the data points are distributed: whether they are homogenous, distributed along a spectrum without clear division, or divisible into one or more subsets.[34] Cluster analysis separates (partitions) data sets into groups (clusters) on the basis of patterns in the data and returns a measurement of the degree to which they are distinct. As an exploratory tool, the success of a cluster model is judged both by its mathematical outcomes and by how interesting the partition is to broader questions in the field of study. As applied to a population of children with neurologic injury, cluster analysis may identify the existence and characteristics of naturally occurring, recognizable subpopulations (clusters). Although cluster analysis cannot be used to specify the cause for the latent structure, the identification of clusters implies that such a cause exists. That cause may be inferred from the nature of the identified data patterns and by comparing the results to that of other studies in which researchers directly address potential causes.

We conducted a retrospective secondary analysis of the combined Pediatric Brain Injury Network (PediBIRN) derivation and validation study data sets (N = 500).[35,36] We hypothesized that cluster analysis, applied to a cohort of young, hospitalized patients with acute neurotrauma, would identify clusters of patients with a discreet clinical presentation. We further hypothesized that one or more clusters would have a relationship to previously described clinical descriptions, historical indications, and physician diagnosis of AHT. Here, we report the results of 3 cluster analysis algorithms, compare and contrast the patients within each cluster, and draw conclusions regarding the relationships between cluster assignment, physicians' diagnosis of AHT, and the concept of the triad.

METHODS

Study Design

We applied 3 different cluster algorithms to sort PediBIRN patients into distinct cohorts. We then compared the frequency of prospectively defined clinical findings across clusters defined by each algorithm and by the presence or absence of the triad. Additionally, we compared the cluster partitions with each other, pairwise, to assess their similarity.

Population

The PediBIRN patient population used for this secondary analysis includes 500 children aged <3 years hospitalized between 2010 and 2013 across 18 PICUs for management of acute, symptomatic, and traumatic closed cranial or intracranial injury identified by computed tomography (CT) scan or MRI from the PediBIRN derivation and validation studies.[35,36] Eligible children were enrolled consecutively by each participating center. Children with preexisting brain abnormalities and children injured in motor vehicle collisions were excluded. All 500 children in the PediBIRN database were included in cluster analysis.

Data

For each qualifying patient, PediBIRN investigators captured prospective data regarding patients' symptoms, clinical and imaging findings, historical features, local diagnostic impression, and the decision to report the case to children's protective services.[35,36] Explanations of terms were incorporated into the data collection tool; for instance, encephalopathy was defined as "clear impairment or loss of consciousness." During the PediBIRN derivation study of 209 patients at 14 sites, a second physician independently entered the same data points for 20% of patients at each individual site.[35] Data used in cluster analysis met the following criteria: the variable possessed sufficient interrater reliability, as evidenced by a κ score >0.6 in the derivation study, and the variable described an objective feature, such as symptoms, examination findings, radiologic findings, and clinical course. Data elements that reflected some

subjective assessment, such as consistency of provided history, final diagnostic impression, or child protective services reporting decisions, were not used in cluster analysis but were analyzed when comparing the resulting clusters.

Cluster Analysis

Throughout the article, we will use the word "cluster" to indicate a group of patients identified by 1 of the 3 cluster algorithms, "partition" as a verb to indicate the process of dividing patients into clusters by 1 of the 3 cluster algorithms, and "partition" as a noun to indicate the structure of clusters across the entire research cohort developed by 1 of the 3 cluster algorithms. The 500 PediBIRN patients were partitioned by using 3 different cluster analysis algorithms: K-means clustering, divisive hierarchical clustering, and agglomerative hierarchical clustering.[34] Cluster analysis was performed in the statistical package R, version 3.6.2, by using the "pam," "diana," and "hclust" commands.[37] Because data included predominantly dichotomous but also continuous data, Gower's method for measuring distance was used for cluster algorithms.[38]

The K-means algorithm requires that the statistician specify the number of desired clusters. K-means clustering was applied, specifying 2 through 10 clusters, producing 9 separate partitions. The resulting 9 partitions were then evaluated by 2 methods, silhouette width and gap statistic, to determine which single K-means partition had the best mathematical characteristics.[34,39] The optimal cluster number was chosen such that silhouette width was maximized and gap statistic was maximized and within 1 SD of the next larger number of clusters. In divisive and agglomerative hierarchical clustering the "complete" or "farthest neighbor" method was used. The divisive and agglomerative clustering algorithms each produce a tree of clusters and subclusters. The level in each tree producing a cluster number matching the optimized K-means partition was chosen for analysis, 1 partition for divisive hierarchical clustering and 1 partition for agglomerative hierarchical clustering.

We also divided the cohort into 2 groups by the presence or absence of the triad. For this process, we defined the triad as the presence of any SDH(s) or fluid collection(s), acute encephalopathy before admission, and RH(s) described by an ophthalmologist as dense, extensive, covering a large surface area and/or extending to the ora serrata (extensive RH). As discussed in the literature, the triad does not distinguish nuances in subdural collections, encephalopathy, or RH. The PediBIRN data set did not collect data on lesser degrees of RH. As such, our operational definition of the triad, within this data set, is more restrictive than the triad discussed in critical literature.

Postclustering Analysis

We performed 2 additional statistical analyses after partition into 2 clusters. For each algorithm's partition, we compared the relative frequencies of clinical variables between the clusters. Significance was determined by χ^2 tests or Fisher's exact tests, as appropriate, in which the Haldane-Anscombe correction was applied if any cell count was 0. Strength of association was determined by odds ratios (ORs) with 95% confidence intervals (CIs). This was done both for variables used to partition the clusters and for select variables not used in cluster analysis.

We also analyzed the relationship between the partitions developed by each algorithm and by the triad. Contingency tables were created to analyze patient sorting into the resulting clusters. Similarity between resulting partitions was assessed for significance by χ^2 analysis and for strength by the accuracy of 1 partition at predicting a comparison partition. Treating physicians' determination of abuse likelihood was dichotomized (definite and/or probable AHT versus undetermined or definite and/or probable non-AHT).

RESULTS

The 9 separate results, created by using the K-means algorithm specifying 2 through 10 clusters, were each evaluated by the silhouette and the gap-statistic methods. The K-means result in which the statistician specified 2 clusters had the best mathematical characteristics. The silhouette width of this solution was 0.22, and the gap statistic was 0.48.

To compare the results of divisive and agglomerative hierarchical analysis to the results of the K-means results, the 2-cluster partitions of the hierarchical algorithms were chosen for further analysis. Silhouette widths and gap statistics are provided here, for description, but were not used in choosing the 2 cluster solutions for these algorithms. The silhouette width for partition into 2 clusters by the divisive hierarchical algorithm was 0.38, and the gap statistic was 0.23. For partition by the agglomerative hierarchical algorithm, silhouette width was 0.35, and the gap statistic was 0.23. Although physicians' final diagnosis of definitive or probable AHT was not used in developing any partition, it was substantially associated ($P < .001$ and OR>10) with 1 of the clusters produced by each algorithm, which we will refer to as cluster 1. We will refer

to the cluster that did not associate with physician-diagnosed AHT as cluster 2.

The triad, by design, divides the population into 2 clusters. The cluster that manifested the triad also associated strongly to physicians' final diagnosis of definitive or probable AHT and will be referred to as cluster 1. We will refer to individual clusters by the partition method and their numerical label (K-means 1, K-means 2, divisive 1, divisive 2, agglomerative 1, agglomerative 2, triad 1, and triad 2)

The within-algorithm comparisons of cluster 1 with cluster 2 (K-means 1 with K-means 2, divisive 1 with divisive 2, agglomerative 1 with agglomerative 2, and triad 1 with triad 2) produced informative results. Most variables occurred with significantly greater frequency ($P < .05$) in 1 of the clusters for each of the 3 mathematical algorithms. (Table 1) Variables making the most substantial contribution to differentiating clusters ($P < .001$ and OR > 10 or <0.1 in all 3 algorithms) were imaging patterns indicating brain hypoxemia-ischemia in any distribution; acute encephalopathy before admission, encephalopathy lasting >24 hours, and encephalopathy lasting >24 hours with subsequent deterioration; respiratory compromise before admission; the presence of SDH or fluid collection; and ophthalmologist-confirmed retinoschisis.

By definition, clinical encephalopathy, SDH or fluid collection, and extensive RH made a substantial contribution ($P < .001$ and OR > 10 or <0.1) to partitioning by the triad. There were differences between partition by the triad and partition by the K-means, divisive, and agglomerative algorithms. Imaging indications of parenchymal brain injury and retinoschisis had lower ORs in partition by the triad than the 3 cluster algorithms. The duration of encephalopathy differed. Brief encephalopathy that resolved before admission associated statistically with the triad-1 cluster but was nonsignificant in the K-means partition and associated with the divisive 2 and agglomerative 2 clusters. Other differences in encephalopathy duration and SDH distribution may be seen in Table 1.

The ages of the children in the K-means 1 cluster of the K-means partition were slightly younger (mean 8.61 vs 10.39 months) than those in the K-means 2 cluster ($P = .046$) (Table 2). There was no statistically significant difference in mean age between divisive 1 and divisive 2, agglomerative 1 and agglomerative 2, and triad 1 and triad 2. The ability to walk or cruise was not statistically different between clusters 1 and 2 in any partition.

When variables not used in the cluster algorithms were analyzed, there were significant differences between clusters 1 and 2, as defined by each of the 3 algorithms and by the triad (Table 3). For each algorithm and for the triad, caregiver admission of abuse, physician-identified changes in the reported clinical history, and physician-identified developmental inconsistencies in reported child behavior were all significantly more frequent in cluster 1. Independently witnessed unintentional injury was significantly less frequent in cluster 1. Caregiver admission of abuse had a substantial ($P < .001$ and OR > 10) relationship to being in the K-means 1 cluster. Independently witnessed unintentional injury had a substantial ($P < .001$ and OR < 0.10) relationship to being in the triad-2 cluster.

Pairwise comparisons of 6 combinations between the 3 algorithms or the triad revealed significant similarities in their partitions, each achieving a $\chi^2 < 0.001$. (Table 4) The K-means and divisive hierarchical algorithms were the most closely related, with an accuracy of 94.8% in predicting one another. All 3 mathematical algorithms agreed in the assignment of 87.8% of patients, 122 to cluster 1 and 317 to cluster 2.

DISCUSSION

The evidence base for diagnosing AHT in children with neurotrauma relies on a series of case-control studies and meta-analyses of those studies.[14–30] Cohorts of AHT cases have been separated from controls by various methods: physician diagnosis,[15,16,22] consensus opinion of a multidisciplinary team,[14,17,18,22,23,26,28,29] predefined criteria designed to avoid indicators under study,[14,19–21,29] and confessed abuse versus a public non-AHT event.[22] Results of the various methods have been consistent, lending strength to the findings. Some authors have pointed to circularity inherent in some of these methods, dismissing the literature as invalid.[31] Cluster analysis uses an entirely different approach, identifying divisions in a body of data by the mathematical distribution of data points. By applying cluster analysis to a cohort of children with neurotrauma, excluding data referencing physician diagnosis or judgements of historical consistency, we have sorted patients without this circularity.[34]

The results of 3 mathematical cluster analyses indicate that the population of children with neurologic injury admitted to a PICU has at least 2 distinguishable subpopulations. As a correlate, these results reject that the objective clinical presentations of these

TABLE 1 Between-Cluster Comparisons of 32 Reliable Clinical and Radiologic Variables Included in the Unsupervised Cluster Analysis, Listed in Descending Order of Their Relative Importance to Patient Assignment into Cluster 1

Variable	K-Means	Divisive Hierarchical	Agglomerative Hierarchical	Triad
Acute encephalopathy before admission, lasting >24 h, with deterioration				
P	<.001[a]	<.001[a]	<.001[a]	<.001[b]
OR (95% CI)	335.30 (46.05–2441.34)[a]	519.17 (70.97–3797.66)[a]	72.15 (32.98–157.87)[a]	9.00 (5.44–14.89)[b]
Bilateral brain hypoxia, ischemia, and/or swelling				
P	<.001[a]	<.001[a]	<.001[a]	<.001[b]
OR (95% CI)	91.21 (42.14–197.43)[a]	141.10 (66.77–298.20)[a]	116.32 (58.90–229.71)[a]	7.76 (4.93–12.24)[b]
Brain hypoxia, ischemia, and/or swelling involving the subcortical brain				
P	<.001[a]	<.001[a]	<.001[a]	<.001[b]
OR (95% CI)	70.44 (32.73–151.62)[a]	111.00 (52.04–236.75)[a]	170.74 (78.85–369.71)[a]	6.93 (4.39–10.94)[b]
Acute encephalopathy before admission, lasting >24 h				
P	<.001[a]	<.001[a]	<.001[a]	<.001[a]
OR (95% CI)	51.00 (28.59–90.99)[a]	68.48 (17.83–96.68)[a]	29.45 (17.33–50.03)[a]	10..35 (6.50–16.47)[a]
Any brain hypoxia, ischemia, and/or swelling				
P	<.001[a]	<.001[a]	<.001[a]	<.001[b]
OR (95% CI)	36.70 (21.83–61.69)[a]	96.89 (49.02–191.5)[a]	144.11 (60.12–345.44)[a]	6.87 (4.39–10.75)[b]
Acute encephalopathy before admission				
P	<.001[a]	<.001[a]	<.001[a]	<.001[a]
OR (95% CI)	38.65 (19.00–78.60)[a]	41.52 (17.83–96.68)[a]	13.20 (7.18–24.27)[a]	381.73 (23.56–6184.68)[a]
Retinoschisis confirmed by an ophthalmologist				
P	<.001[a]	<.001[a]	<.001[a]	<.001[b]
OR (95% CI)	30.76 (7.23–130.85)[a]	40.38 (9.48–172.04)[a]	16.76 (6.27–44.83)[a]	8.42 (3.74–18.94)[b]
Acute respiratory compromise before admission				
P	<.001[a]	<.001[a]	<.001[a]	<.001[a]
OR (95% CI)	22.81 (14.10–36.89)[a]	35.30 (20.05–62.14)[a]	19.44 (11.46–32.98)[a]	10.55 (6.49–17.16)[a]
Any SDH(s) or fluid collection(s)				
P	<.001[a]	<.001[a]	<.001[a]	<.001[a]
OR (95% CI)	21.88 (9.96–48.06)[a]	16.04 (7.30–35.26)[a]	11.30 (5.37–23.78)[a]	186.25 (11.50–3017.02)[a]
Interhemispheric SDH				
P	<.001[a]	<.001[a]	<.001[b]	<.001[a]
OR (95% CI)	18.95 (11.88–30.25)[a]	13.01 (8.17–20.73)[a]	9.30 (5.86–14.76)[b]	11.47 (6.98–18.85)[a]
RH(s) described by an ophthalmologist as dense, extensive, covering a large surface area, and/or extending to the ora serrata				
P	<.001[a]	<.001[a]	<.001[b]	<.001[a]
OR (95% CI)	17.85 (11.07–28.79)[a]	11.05 (7.06–17.30)[a]	9.62 (6.13–15.11)[b]	3148.13 (190.59–52 000.93)[a]
Bilateral SDH				
P	<.001[a]	<.001[b]	<.001[b]	<.001[b]
OR (95% CI)	13.02 (8.38–20.24)[a]	8.36 (5.41–12.90)[b]	6.90 (4.44–10.72)[b]	5.87 (3.77–9.16)[b]

TABLE 1 Continued

Variable	K-Means	Divisive Hierarchical	Agglomerative Hierarchical	Triad
Seizure(s) before admission				
P	<.001[a]	<.001[b]	<.001[b]	<.001[b]
OR (95% CI)	12.89 (8.23–20.19)[a]	7.89 (5.13–12.13)[b]	4.56 (2.99–6.95)[b]	7.79 (4.96–12.23)[b]
Brain contusion(s), laceration(s) or hemorrhage(s) compatible with diffuse traumatic axonal injury				
P	<.001[b]	<.001[b]	<.001[a]	.343[c]
OR (95% CI)	8.63 (2.86–26.09)[b]	8.32 (2.99–23.18)[b]	30.42 (6.98–132.57)[a]	1.56 (0.62–3.97)[c]
AST or ALT level >80 IU/L any time after hospital admission				
P	<.001[b]	<.001[b]	<.001[b]	<.001[b]
OR (95% CI)	5.90 (3.69–9.44)[b]	7.15 (4.46–11.47)[b]	5.53 (3.48–8.80)[b]	3.75 (2.36–5.94)[b]
Skeletal survey that revealed fracture(s) moderately or highly specific for abuse				
P	<.001[b]	<.001[b]	<.001[b]	<.001[b]
OR (95% CI)	5.40 (3.28–8.88)[b]	4.29 (2.65–6.94)[b]	4.20 (2.59–6.81)[b]	2.99 (1.84–4.87)[b]
Any brain parenchymal contusion(s), laceration(s), or hemorrhage(s) involving the subcortical (or deeper) brain				
P	<.001[b]	<.001[b]	<.001[a]	.096[c]
OR (95% CI)	5.50 (2.58–11.69)[b]	6.23 (2.98–13.04)[b]	14.34 (6.11–33.68)[a]	1.82 (0.89–3.71)[c]
Any bruising involving the child's ear(s), neck, or torso				
P	<.001[b]	<.001[b]	<.001[b]	<.001[b]
OR (95% CI)	3.42 (2.19–5.33)[b]	4.12 (2.63–6.47)[b]	3.80 (2.42–5.99)[b]	4.24 (2.67–6.71)[b]
CT scan–confirmed intraabdominal injuries				
P	.004[b]	.001[b]	.01[b]	1.000[c]
OR (95% CI)	4.29 (1.47–12.56)[b]	5.52 (1.88–16.17)[b]	3.74 (1.36–10.23)[b]	1.02 (0.32–3.23)[c]
Skin bruising, abrasion(s), or laceration(s) in ≥2 distinct locations other than the knees, shins, or elbows				
P	<.001[b]	<.001[b]	<.001[b]	<.001[b]
OR (95% CI)	2.71 (1.73–4.23)[b]	3.04 (1.93–4.78)[b]	2.20 (1.39–3.49)[b]	2.41 (1.51–3.84)[b]
Any subarachnoid hemorrhage(s)				
P	<.001[b]	<.001[b]	<.001	.017[b]
OR (95% CI)	2.42 (1.63–3.58)[b]	3.31 (2.20–4.97)[b]	3.52 (2.32–5.35)[b]	1.68 (1.10–2.57)[b]
Any brain parenchymal contusion(s), laceration(s), or hemorrhage(s)				
P	.008[b]	<.001[b]	<.001[b]	.671[c]
OR (95% CI)	1.83 (1.17–2.86)[b]	2.26 (1.43–3.55)[b]	3.13 (1.98–4.97)[b]	1.12 (0.68–1.85)[c]
Acute encephalopathy before admission, resolved before admission				
P	.442[c]	<.001[d]	<.001[d]	<.001[b]

TABLE 1 Continued

Variable	K-Means	Divisive Hierarchical	Agglomerative Hierarchical	Triad
OR (95% CI)	0.80 (0.45–10.42)[c]	0.27 (0.12–0.60)[d]	0.37 (0.17–0.80)[d]	2.54 (1.46–4.41)[b]
Child cruising or walking before admission				
P	.06[c]	.417[c]	.159[c]	.572[c]
OR (95% CI)	0.67 (0.44–1.02)[c]	0.84 (0.54–1.29)[c]	0.72 (0.45–1.14)[c]	0.88 (0.55–1.39)[c]
Unilateral SDH				
P	.007[d]	.078[c]	.115[c]	.863[c]
OR (95% CI)	0.54 (0.34–0.85)[d]	0.66 (0.42–1.05)[c]	0.68 (0.42–1.10)[c]	1.04 (0.66–1.66)[c]
Acute encephalopathy before admission, resolved within 24 h				
P	.079[c]	.032[d]	.001[d]	.507[c]
OR (95% CI)	0.57 (0.30–1.07)[c]	0.47 (0.23–0.95)[d]	0.24 (0.09–0.61)[d]	0.80 (0.41–1.56)[c]
Craniofacial bruising, abrasion(s), subgaleal hematoma(s), or cephalohematoma(s)				
P	<.001[d]	.024[d]	.091[d]	.003[d]
OR (95% CI)	0.50 (0.34–0.73)[d]	0.64 (0.43–0.94)[d]	0.71 (0.47–1.06)[d]	0.54 (0.36–0.81)[d]
Any skull fracture(s) other than an isolated, nondiastatic, linear parietal skull fracture				
P	.002[d]	.071[c]	.689[c]	<.001[d]
OR (95% CI)	0.51 (0.33–0.79)[d]	0.67 (0.43–1.04)[c]	0.91 (0.59–1.42)[c]	0.27 (0.15–0.49)[d]
Any skull fracture(s)				
P	<.001[d]	<.001[d]	<.001[d]	<.001[d]
OR (95% CI)	0.15 (0.10–0.23)[d]	0.23 (0.15–0.34)[d]	0.33 (0.21–0.50)[d]	0.13 (0.08–0.21)[d]
An isolated, unilateral, nondiastatic, linear parietal skull fracture				
P	<.001[e]	<.001[d]	<.001[d]	<.001[d]
OR (95% CI)	0.08 (0.04–0.18)[e]	0.11 (0.05–0.23)[d]	0.13 (0.06–0.28)[d]	0.14 (0.06–0.32)[d]
Any epidural hemorrhage(s)				
P	<.001[e]	<.001[e]	<.001[d]	<.001[d]
OR (95% CI)	0.07 (0.03–0.21)[e]	0.10 (0.03–0.27)[e]	0.11 (0.04–0.32)[d]	0.13 (0.05–0.36)[d]

ALT, alanine aminotransferase; AST, aspartate aminotransferase.
[a] Substantial association with cluster 1 ($P < .001$ and OR > 10).
[b] Significant association with cluster 1 ($P < .001$ and OR ≤ 10 but >1).
[c] Statistically insignificant ($P > .05$).
[d] Significant association with cluster 2 ($P < .001$ and OR ≥ 0.10 but <1).
[e] Substantial association with cluster 2 ($P < .001$ and OR < 0.10).

children occur across a spectrum of severity, without clear division, or in random combinations of findings without pattern. Cluster results are generally considered appropriate when silhouette width is >0.2 and are considered strong when silhouette width is >0.6.[34] Each of the 3 algorithms we deployed generated 2-cluster results with a silhouette width >0.2, each partition pair correlated with a χ^2 <0.001 and accuracy >88%, and all agreed on the assignment of 87.8% of patients, indicating that partition was justifiable and robust. Although partition implies that a latent factor operated to produce 2 subpopulations

TABLE 2 Age Differences Between Clusters for 3 Algorithms and the Triad

	K-Means		Divisive		Agglomerative		Triad	
Age, mo	K-means 1	K-means 2	Divisive 1	Divisive 2	Agglomerative 1	Agglomerative 2	Triad 1	Triad 2
Mean	8.61	10.39	9.36	9.94	9.04	10.04	8.35	10.23
SD	98.94	9.75	9.39	9.56	9.45	9.52	7.74	9.98
Difference, *P*	.046		.533		.299		.057	

Significance tested by unpaired *t* test.

TABLE 3 Between-Cluster Comparisons of Variables Not Included in the Unsupervised Cluster Analysis, Listed in Descending Order of Their Relative Importance to Patient Assignment into the Cluster 1

	K-Means	Divisive Hierarchical	Agglomerative Hierarchical	Triad
Physician final diagnosis of definitive or probable AHT				
P	<.001[a]	<.001[a]	<.001[a]	<.001[a]
OR (95% CI)	38.97 (20.62–73.66)[a]	27.85 (14.48–53.59)[a]	19.29 (10.26–36.25)[a]	45.37 (18.08–113.83)[a]
Caregiver admission of AHT				
P	<.001[a]	<.001[b]	<.001[b]	<.001[b]
OR (95% CI)	11.68 (4.41–30.92)[a]	5.91 (2.72–12.82)[b]	6.07 (2.84–12.97)[b]	6.85 (3.20–14.67)[b]
History of unintentional trauma consistent with repetition over time				
P	<.001[c]	<.001[c]	<.001[c]	<.001[c]
OR (95% CI)	0.10 (0.07–0.17)[c]	0.12 (0.08–0.20)[c]	0.14 (0.09–0.23)[c]	0.20 (0.12–0.32)[c]
History of AHT consistent with the child's gross motor skills				
P	<.001[c]	<.001[c]	<.001[c]	<.001[c]
OR (95% CI)	0.16 (0.11–0.25)[c]	0.20 (0.13–0.31)[c]	0.21 (0.13–0.33)[c]	0.31 (0.20–0.47)[c]
Independently witnessed AHT				
P	<.001[c]	<.001[c]	.002[c]	<.001[d]
OR (95% CI)	0.12 (0.04–0.39)[c]	0.15 (0.05–0.50)[c]	0.18 (0.05–0.59)[c]	0.06 (0.009–0.46)[d]

[a] Substantial association with cluster 1 ($P < .001$ and OR > 10).
[b] Significant association with cluster 1 ($P < .001$ and OR ≤ 10 but >1).
[c] Significant association with cluster 2 ($P < .001$ and OR ≥ 0.10 but <1).
[d] Substantial association with cluster 2 ($P < .001$ and OR < 0.10).

within the PediBIRN cohort, our cluster results alone do not reveal that the difference was AHT.

For each algorithm, children in cluster 1 manifested significantly increased rates of SDH other than unilateral contact SDH, acute encephalopathy that did not resolve within 24 hours, extensive RH, retinoschisis, abuse-associated fractures on skeletal survey, concerning traumatic skin findings, and laboratory or imaging indications of intraabdominal injury. By contrast, children in cluster 2 manifested high rates of findings that indicate impact to the head: skull fracture, epidural hemorrhage, and extracranial craniofacial injury. It bears repeating that partition into clusters 1 and 2 made no reference to physicians' diagnosis of AHT or any research definition of AHT. And yet, mathematical partition based on objective, reliably determined clinical variables replicated the results of decades of literature on the characteristics of children with AHT.[14–30] By replicating the segregation of clinical findings in the case-control literature, the cluster analyses demonstrate that the division of cases from controls reflected natural latent divisions in the larger population of young children with neurotrauma. The fact that these differences extended beyond neurologic and eye findings to include skeletal fractures, skin injuries, and visceral injuries supports the inference that inflicted trauma was closely related to the latent variable responsible for partition.

Given that partition into clusters 1 and 2 replicates clinical associations found by researchers into the characteristics of children with AHT, it is unsurprising that partition was significantly associated with clinical diagnosis of AHT by treating physicians. We suspect that diagnosing physicians applied the very case-control studies we have previously referenced in their clinical determinations. Physicians also appear to have considered nuances and findings not used by the clustering algorithms when making their diagnoses, diagnosing both AHT and non-AHT within clusters 1 and 2 of each algorithm. This is consistent with the finding that silhouette width never

TABLE 4 Agreement Between the Various Methods of Partition Into Cluster 1 and Cluster 2

	K-Means	Divisive Hierarchical	Agglomerative Hierarchical
Divisive hierarchical			
P	<.001	—	—
Accuracy of agreement between partitions	0.948	—	—
95% CI	0.928–0.966	—	—
Agglomerative hierarchical			
P	<.001	<.001	—
Accuracy of agreement between partitions	0.884	0.924	—
95% CI	0.853–0.911	0.897–0.946	—
Triad			
P	<.001	<.001	<.001
Accuracy of agreement between partitions	0.836	0.824	0.800
95% CI	0.801–0.867	0.788 –0.856	0.762–0.834

Rows indicate *P* value, accuracy of agreement between partitions, and 95% CIs for accuracy. —, not applicable.

exceeded 0.6 for any cluster algorithm, indicating that the 2 clusters are not highly discreet.

Authors who have questioned the existing literature guiding diagnosis of AHT have also suggested that research should rely on confession to identify children with AHT.[31] We treated both confessions and the observations of independent witnesses as subjective and not susceptible to reliability testing, excluding consideration of these variables by the cluster algorithms. Despite this, when there was an admission of AHT, it associated with cluster 1, and when there was an independent witness to an unintentional injury, it associated with cluster 2. Viewed another way, nonmedical, firsthand witness data agreed with cluster partition and with AHT characteristics identified in decades of literature and in our 3 mathematical partitions.

Final, interesting outcomes were age and developmental differences between clusters. Much earlier research into characteristics of AHT have found children with AHT to be significantly younger than those in comparison groups.[14,17–19,22,23,25–28,30] Only 1 of the clusters, K-means 1, had a significant association with age, based on a difference of 1.78 months, with a *P* of .046. The ability to walk or cruise was not statistically different between clusters 1 and 2 of any algorithm. It follows that physiologic and developmental variables related to age and maturity are unlikely to be responsible for other differences between cluster pairs.

We began this study to see whether a noncircular research method, blind to issues of abuse, would replicate the clinical associations found in the case-control literature that references a determination of abuse in its methods. What we found was not only that it did but that the clinical findings making the most contribution to partition were encephalopathy, SDH, and retinal injury. This immediately called to mind the triad of findings that some authors have asserted are both poorly supported by research literature and heavily relied on by physicians to diagnose AHT.[31–33] For this reason, we looked at partition by the closest surrogate we could construct for this triad. The partition by these criteria was statistically similar to the partitions derived by the 3 mathematical algorithms, with sorting accuracies of 80.0% to 83.6%. Furthermore, many of the associations with extracephalic findings attributable to abuse, and with available firsthand reports of abuse or unintentional injury, were preserved in this partition. As such, it is unsurprising that the triad, when incorporating the additional component of requiring that RHs be "dense, extensive and covering a large surface area or extending to the ora serrata," has a high specificity for physicians' diagnosis of AHT (98.0%), although a modest sensitivity (47.8%) and accuracy (73.2%). The triad, too, appears to reflect a natural division in the population of children with neurologic injury, reflective of a latent variable that is related to the diagnosis of AHT.

We have identified several limitations to this study. The data used in cluster analysis were captured in a PICU setting on patients with acute symptomatic injury. It is unknown if the results would have been the same in a non-PICU setting or in children with nonacute intracranial injury. Bias may be introduced in ascertainment of certain findings. All patients had a history, physical examination, and head imaging by CT or MRI. Retinal examination and skeletal survey were performed at physicians' discretion, and thus the absence of fractures and RHs could have been the result of omitted rather than normal ophthalmologic or radiologic examinations. Among the 500 patients, 322 patients underwent both skeletal survey and retinal examination, and 109 patients underwent neither evaluation. Retinoschisis was the only inconsistently ascertained finding that had a substantial ($P < .001$ and OR > 10) association with cluster 1 and only occurred in 30 cases. Thus, the variables that most substantially contributed to partition were universally assessed, and the influence of ascertainment bias appears to be minimized. Finally, the variables captured in the study were chosen as relevant to child abuse. It is conceivable that the inclusion of other variables might have produced different results.

CONCLUSIONS

When mathematical clustering methods based solely on objectively and reliably ascertained clinical variables were used to analyze a cohort of young children with neurologic injury, 2 subpopulations emerged. One of these subpopulations was characterized by an elevated prevalence of imaging patterns indicating brain hypoxemia-ischemia, SDH other than unilateral contact SDH, retinal findings, extracephalic injuries, the absence of clear evidence of head impact, caregiver admission of AHT, and physician-identified inconsistencies in provided trauma histories. These differences replicate differences previously found in case-control literature as distinguishing AHT, differences that treating physicians appeared to view as highly informative to the diagnosis of AHT. We conclude that there are aspects unique to AHT that produce discernible subpopulations and that the division of cases from controls

in previous research on AHT reflected a natural division in the larger cohort. Partition of these same patients by the presence or absence of a proposed triad of findings produces similar results and is likely related to the same latent factor. These results support the preponderant diagnostic practices in the pediatric community. By arriving at similar results through very different methods, this study validates previous literature and should strengthen physician's confidence in the current diagnostic paradigm and their presentation of that paradigm in court.

ACKNOWLEDGMENTS

The authors acknowledge and thank the remaining PediBIRN investigators who helped to capture the data used in this secondary analysis: Bruce E. Herman, MD, and Antoinette Laskey, MD, MPH (Primary Children's Medical Center, Salt Lake City, UT); Douglas F. Willson, MD, and Robin Foster, MD (The Children's Hospital of Richmond, Richmond, VA); Veronica Armijo-Garcia, MD, and Sandeep K. Narang, MD, JD (University of Texas Health Sciences Center at San Antonio, San Antonio, TX); Christopher Carroll, MD (Connecticut Children's Medical Center, Hartford, CT); Deborah A. Pullin, BSN, APRN (Dartmouth-Hitchcock Medical Center, Lebanon, NH); Jeanine M. Graf, MD, and Reena Isaac, MD (Texas Children's Hospital, Houston, TX); Terra N. Frazier, DO, and Kelly Tieves, MD (Children's Mercy Hospital, Kansas City, MO); Edward Truemper, MD, and Suzanne Haney, MD (Children's Hospital of Omaha, Omaha, NE); Kerri Weeks, MD, and Lindall E. Smith, MD (Wesley Medical Center, Wichita, KS); Renee A. Higgerson, MD, and George A. Edwards, MD (Dell Children's Medical Center of Central Texas, Austin, TX); Nancy S. Harper, MD, FAAP and Karl L. Serrao, MD, FAAP, FCCM (Driscoll Children's Hospital, Corpus Christi, TX); Andrew Sirotnak, MD, Joseph Albietz, MD, and Antonia Chiesa, MD (Children's Hospital Colorado, Denver, CO); Christine McKiernan, MD (Baystate Medical Center, Springfield, MA); Mark S. Dias, MD (Penn State College of Medicine, Hershey, PA); Michael Stoiko, MD, Debra Simms, MD, FAAP, and Sarah J. Brown, DO, FACOP, FAAP (Helen DeVos Children's Hospital, Grand Rapids, MI); Amy Ornstein, MD, FRCPC (IWK Health Centre, Halifax, Nova Scotia); and Phil Hyden, MD (Children's Hospital of Central California, Madera, CA)

ABBREVIATIONS

AHT: abusive head trauma
CI: confidence interval
CT: computed tomography
OR: odds ratio
PediBIRN: Pediatric Brain Injury Research Network
RH: retinal hemorrhage
SDH: subdural hemorrhage

PEDIATRICS (ISSN Numbers: Print, 0031-4005; Online, 1098-4275).

Address correspondence to Stephen C. Boos, MD, Baystate Medical Center, Family Advocacy Center, 300 Carew St, Springfield, MA 01199. E-mail: stephen.boosmd@bhs.org

PEDIATRICS (ISSN Numbers: Print, 0031-4005; Online, 1098-4275).

FINANCIAL DISCLOSURE: The authors have indicated they have no financial relationships relevant to this article to disclose.

FUNDING: Supported by Dartmouth-Hitchcock Medical Center, a private family foundation, The Gerber Foundation, Penn State University, Penn State Health Milton S. Hershey Medical Center, and the National Institutes of Health (grant P50HD089922). These funding organizations had no role in the study design; in the collection, analysis, and interpretation of the data; in the writing of the report; and in the decision to submit the article for publication. The content of this study is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health. Funded by the National Institutes of Health (NIH).

POTENTIAL CONFLICT OF INTEREST: The authors have indicated they no potential conflicts of interest to disclose.

COMPANION PAPER: A companion to this article can be found online at www.pediatrics.org/cgi/doi/10.1542/peds.2021-054009.

REFERENCES

1. Badger S, Waugh MC, Hancock J, Marks S, Oakley K. Short term outcomes of children with abusive head trauma two years post injury: a retrospective study. *J Pediatr Rehab Med.* 2020;13(3):241–253
2. Barlow KM, Minns RA. Annual incidence of shaken impact syndrome in young children. *Lancet.* 2000;356(9241):1571–1572
3. Barlow K, Thompson E, Johnson D, Minns RA. The neurological outcome of non-accidental head injury. *Pediatr Rehabil.* 2004;7(3):195–203
4. Barlow KM, Thomson E, Johnson D, Minns RA. Late neurologic and cognitive sequelae of inflicted traumatic brain injury in infancy. *Pediatrics.* 2005;116(2) Available at: www.pediatrics.org/cgi/content/full/116/2/e174
5. Chevignard MP, Lind K. Long-term outcome of abusive head trauma.

Pediatr Radiol. 2014;44(Suppl 4): S548–S558

6. Eismann EA, Theuerling J, Cassedy A, Curry PA, Colliers T, Makoroff KL. Early developmental, behavioral, and quality of life outcomes following abusive head trauma in infants. *Child Abuse Negl*. 2020;108:104643
7. Fanconi M, Lips U. Shaken baby syndrome in Switzerland: results of a prospective follow-up study, 2002-2007. *Eur J Pediatr*. 2010;169(8):1023–1028
8. Keenan HT, Runyan DK, Marshall SW, Nocera MA, Merten DF, Sinal SH. A population-based study of inflicted traumatic brain injury in young children. *JAMA*. 2003;290(5):621–626
9. Lind K, Toure H, Brugel D, Meyer P, Laurent-Vannier A, Chevignard M. Extended follow-up of neurological, cognitive, behavioral and academic outcomes after severe abusive head trauma. *Child Abuse Negl*. 2016;51:358–367
10. Makaroff KL, Putnam FW. Outcomes of infants and children with inflicted traumatic brain injury. *Dev Med Child Neurol*. 2003;45(7):497–502
11. Nuño M, Ugiliweneza B, Zepeda V, et al. Long-term impact of abusive head trauma in young children. *Child Abuse Negl*. 2018;85:39–46
12. Rebbe R, Mienko JA, Martinson ML. Incidence and risk factors for abusive head trauma: a population-based study. *Child Abuse Rev*. 2020;29(3):195–207
13. Talvik I, Metsvaht T, Leito K, et al. Inflicted traumatic brain injury (ITBI) or shaken baby syndrome (SBS) in Estonia. *Acta Paediatr*. 2006;95(7):799–804
14. Amagasa S, Matsui H, Tsuji S, Uematsu S, Moriya T, Kinoshita K. Characteristics distinguishing abusive head trauma from accidental head trauma in infants with traumatic intracranial hemorrhage in Japan. *Acute Med Surg*. 2018;5(3): 265–271
15. Bechtel K, Stoessel K, Leventhal JM, et al. Characteristics that distinguish accidental from abusive injury in hospitalized young children with head trauma. *Pediatrics*. 2004;114(1):165–168
16. Binenbaum G, Mirza-George N, Christian CW, Forbes BJ. Odds of abuse associated with retinal hemorrhages in children suspected of child abuse. *J AAPOS*. 2009;13(3):268–272
17. Burns J, Rohl S, Marth D, Proctor D, Amin R, Sekhon C. Which clinical features of children on initial presentation to the emergency department with head injury are associated with clinically important traumatic brain injury, classification as abuse, and poor prognosis [published online ahead of print September 10, 2020]? *Pediatr Emerg Care*. doi:10.1097/PEC.0000000000002239
18. Ewing-Cobbs L, Prasad M, Kramer L, et al. Acute neuroradiologic findings in young children with inflicted or noninflicted traumatic brain injury. *Childs Nerv Syst*. 2000;16(1):25–33
19. Feldman KW, Bethel R, Shugerman RP, Grossman DC, Grady MS, Ellenbogen RG. The cause of infant and toddler subdural hemorrhage: a prospective study. *Pediatrics*. 2001;108(3):636–646
20. Hymel KP, Makoroff KL, Laskey AL, Conaway MR, Blackman JA. Mechanisms, clinical presentations, injuries, and outcomes from inflicted versus noninflicted head trauma during infancy: results of a prospective, multicentered, comparative study. *Pediatrics*. 2007; 119(5):922–929
21. Hymel KP, Wang M, Chinchilli VM, et al; Pediatric Brain Injury Research Network (PediBIRN) Investigators. Estimating the probability of abusive head trauma after abuse evaluation. *Child Abuse Negl*. 2019;88:266–274
22. Keenan HT, Runyan DK, Marshall SW, Nocera MA, Merten DF. A population-based comparison of clinical and outcome characteristics of young children with serious inflicted and noninflicted traumatic brain injury. *Pediatrics*. 2004;114(3):633–639
23. Kelly P, John S, Vincent AL, Reed P. Abusive head trauma and accidental head injury: a 20-year comparative study of referrals to a hospital child protection team. *Arch Dis Child*. 2015;100(12):1123–1130
24. Kemp AM, Jaspan T, Griffiths J, et al. Neuroimaging: what neuroradiological features distinguish abusive from non-abusive head trauma? A systematic review. *Arch Dis Child*. 2011;96(12): 1103–1112
25. Maguire SA, Kemp AM, Lumb RC, Farewell DM. Estimating the probability of abusive head trauma: a pooled analysis. *Pediatrics*. 2011;128(3). Available at: www.pediatrics.org/cgi/content/full/128/3/e550
26. Morgan LA, Fouzdar Jain S, Svec A, et al. Clinical comparison of ocular and systemic findings in diagnosed cases of abusive and non-abusive head trauma. *Clin Ophthalmol*. 2018;12:1505–1510
27. Piteau SJ, Ward MG, Barrowman NJ, Plint AC. Clinical and radiographic characteristics associated with abusive and nonabusive head trauma: a systematic review. *Pediatrics*. 2012;130(2):315–323
28. Reece RM, Sege R. Childhood head injuries: accidental or inflicted? *Arch Pediatr Adolesc Med*. 2000;154(1):11–15
29. Vinchon M, Defoort-Dhellemmes S, Desurmont M, Dhellemmes P. Accidental and nonaccidental head injuries in infants: a prospective study. *J Neurosurg*. 2005;102(4 Suppl):380–384
30. Vinchon M, de Foort-Dhellemmes S, Desurmont M, Delestret I. Confessed abuse versus witnessed accidents in infants: comparison of clinical, radiological, and ophthalmological data in corroborated cases. *Childs Nerv Syst*. 2010;26(5):637–645
31. Lynøe N, Elinder G, Hallberg B, Rosén M, Sundgren P, Eriksson A. Insufficient evidence for 'shaken baby syndrome' - a systematic review. *Acta Paediatr*. 2017;106(7):1021–1027
32. Lynoe N, Eriksson A. The unspoken shaken baby lie detector algorithm—an analysis of diagnostic procedures in cases of allegedly abusive head trauma without external signs of trauma. *J Res Philosophy History*. 2020;3:52–65
33. Swedish Council on Health Technology Assessment. *Traumatic Shaking: The Role of the Triad in Medical Investigations of Suspected Traumatic Shaking: A Systematic Review*. Stockholm, Sweeden: Swedish Council on Health Technology Assessment; 2016
34. Everitt BS, Landau S, Leese M, Stahl D. *Cluster Analysis*. 5th ed. New York, NY: John Wiley & Sons, Ltd; 2011
35. Hymel KP, Willson DF, Boos SC, et al; Pediatric Brain Injury Research Network (PediBIRN) Investigators. Derivation of a

clinical prediction rule for pediatric abusive head trauma. *Pediatr Crit Care Med*. 2013;14(2):210–220

36. Hymel KP, Armijo-Garcia V, Foster R, et al; Pediatric Brain Injury Research Network (PediBIRN) Investigators. Validation of a clinical prediction rule for pediatric abusive head trauma. *Pediatrics*. 2014;134(6). Available at: www.pediatrics.org/cgi/content/full/134/6/e1537
37. The R Foundation. The R project for statistical computing. Available at: https://www.r-project.org/. Accessed August 20, 2021
38. The R Foundation. gower.dist: computes the gower's distance. Available at: https://www.rdocumentation.org/packages/StatMatch/versions/1.4.0/topics/gower.dist. Accessed August 20, 2021
39. Data Novia. Determining the optimal number of clusters: 3 must know methods. Available at: https://www.datanovia.com/en/lessons/determining-the-optimal-number-of-clusters-3-must-know-methods/. Accessed August 20, 2021