

# Estimating the Probability of Abusive Head Trauma: A Pooled Analysis

**WHAT'S KNOWN ON THIS SUBJECT:** In recent debate in medicolegal circles, clinical basis for a diagnosis of abusive head trauma (AHT) has been questioned. Studies have been underpowered to address the key clinical questions. Estimates of the association of single clinical variables with AHT are recognized.

**WHAT THIS STUDY ADDS:** This study provides an estimate of probability of AHT based on 6 clinical features. In a child younger than 3 who has an intracranial injury and  $\geq 3$  of these features, the positive predictive value of AHT is  $>85\%$  (odds ratio:  $>100$ ).

**AUTHORS:** Sabine Ann Maguire, MRCPI, MRCPCh, FRCPI,<sup>a</sup> Alison Mary Kemp, MRCP, DCH, FRCPCH, FRCP,<sup>a</sup> Rebecca Caroline Lumb, BA,<sup>a</sup> and Daniel Mark Farewell, MMath, PhD<sup>b</sup>

*Departments of <sup>a</sup>Child Health and <sup>b</sup>Primary Care and Public Health, School of Medicine, Cardiff University, Cardiff, United Kingdom*

## KEY WORDS

abusive head trauma, child abuse, diagnostic tool, pooled estimates of probability of AHT

## ABBREVIATIONS

AHT—abusive head trauma  
ICI—intracranial injury  
nAHT—nonabusive head trauma  
PPV—positive predictive value  
OR—odds ratio  
RH—retinal hemorrhage  
CI—confidence interval

Dr Maguire had the initial study idea and was principle investigator for primary systematic review, drafted and revised the article, and gave final approval for publication; Ms Lumb entered all anonymized data into the database, formatted the original article, communicated with all primary authors, and edited the article; Dr Farewell designed and conducted all statistical analyses, wrote the statistical component of article, and contributed to overall editing; and Prof Kemp was program director for primary systematic review and contributed to the conception of analysis, drafts of the article, and subsequent editing.

[www.pediatrics.org/cgi/doi/10.1542/peds.2010-2949](http://www.pediatrics.org/cgi/doi/10.1542/peds.2010-2949)

doi:10.1542/peds.2010-2949

Accepted for publication May 5, 2011

Address correspondence to Sabine Ann Maguire, MRCPI, MRCPCh, FRCPI Wales College of Medicine, Cardiff University, UHW, Heath Site, Cardiff CF14 4XN, United Kingdom. E-mail: sabinemaguire@yahoo.co.uk

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

Copyright © 2011 by the American Academy of Pediatrics

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

## abstract

**CONTEXT AND OBJECTIVE:** To determine which combinations of clinical features assist in distinguishing abusive head trauma (AHT) from nonabusive head trauma.

**METHODS:** Individual patient data from 6 comparative studies of children younger than 3 years with intracranial injury were analyzed to determine the association between AHT and combinations of apnea; retinal hemorrhage; rib, skull, and long-bone fractures; seizures; and head and/or neck bruising. An aggregate analysis of data from these studies used multiple imputation of combined clinical features using a bespoke hotdeck imputation strategy, which accounted for uncertainty arising from missing information.

**RESULTS:** Analyzing 1053 children (348 had AHT), excluding nonsignificant variables (gender, age, skull fractures), for a child with an intracranial injury and 1 or 2 of the 6 features, the positive predictive value (PPV) of AHT varies from 4% to 97% according to the different combinations. Although rarely recorded, apnea is significantly associated with AHT (odds ratio [OR]: 6.89 [confidence interval: 2.08–22.86]). When rib fracture or retinal hemorrhage was present with any 1 of the other features, the OR for AHT is  $>100$  (PPV  $> 85\%$ ). Any combination of 3 or more of the 6 significant features yielded an OR of  $>100$  (PPV for AHT  $> 85\%$ ).

**CONCLUSIONS:** Probabilities of AHT can be estimated on the basis of different combinations of clinical features. The model could be further developed in a prospective large-scale study, with an expanded clinical data set, to contribute to a more refined tool to inform clinical decisions about the likelihood of AHT. *Pediatrics* 2011;128:e550–e564

Abusive head trauma (AHT) remains the commonest form of fatal child abuse and predominantly affects infants.<sup>1</sup> These children classically present with an intracranial injury (ICI) that is evident on neuroimaging or at post mortem, where there is no explanation of trauma, or a history that is inconsistent with the severity of injury or the developmental level of the child. The challenge for the clinician is to determine, firstly, which children should be investigated to exclude AHT; secondly, when concerning features are found, how confident one can be that AHT is the cause rather than nonabusive head trauma (nAHT) or an organic disease. After a full clinical assessment, the multidisciplinary team must piece together all available (eg, clinical, social, forensic) information and determine the likelihood of an abusive etiology. These decisions are difficult, and the pressures to make the correct diagnosis are considerable. A child with missed AHT may later present with a fatal injury<sup>2</sup>; on the other hand, an incorrect diagnosis of AHT will have devastating consequences for the child and family.

When cases of suspected child abuse enter the legal child protection process, the medical evidence contributes heavily to decisions about the future risks to the child, and child care proceedings, regardless of whether the evidence informs a criminal prosecution. In a recent law review article, Turkheimer questioned the diagnosis of AHT on the basis of combinations of clinical features.<sup>3</sup> This stimulated controversial argument in The New York Times: Tuerkheimer suggested that experts are questioning the scientific basis for shaken baby syndrome (sic),<sup>4</sup> which was roundly refuted by responses from M. Barr and J. Leventhal,<sup>5</sup> among others, thus highlighting the need for an agreed scientific basis to contribute to this diagnosis. In

their role as expert witnesses in court, clinicians in the United Kingdom have also come under pressure to support their opinion with scientific evidence, as highlighted by Baroness Kennedy: “A doctor can be convinced based on his or her experience that a defendant is guilty, but unless there is compelling evidence, supported scientifically, he or she should not express that view in criminal proceedings.”<sup>6</sup> As such, an evidence-based calculation that assimilates the clinical data and contributes to determining the probability of AHT would be a valuable adjunct for the professionals concerned.

Although many studies have evaluated the diagnostic utility of individual features in identifying children with AHT, only a small number have explored how combined features predict the condition<sup>7,8</sup>; given that AHT is relatively uncommon in any given population,<sup>9</sup> individual studies are frequently underpowered to draw statistically valid conclusions for multiple combinations of features.<sup>10</sup> Previously, we conducted a meta-analysis of high quality studies that derives the positive predictive values (PPV) and odds ratios (ORs) for AHT for individual features.<sup>11</sup> However, in a clinical scenario, the diagnosis of AHT is based on an interpretation of the combined findings. In this study we aim to combine individual patient data from 6 large population-based comparative studies of AHT and nAHT to estimate the probability of AHT, given different combinations of features identifiable during the initial evaluation of such cases. In so doing, it may provide a statistical estimate to assist clinicians.

## METHODS

We conducted a systematic review of the clinical features of AHT and nAHT<sup>11</sup> (apnea; retinal hemorrhages [RHs]; rib, skull, and long-bone fractures; seizures; and head and/or neck bruising)

and identified 14 high quality comparative studies. We contacted the authors of the most recent larger studies that encompassed the majority of the clinical indicators of AHT and nAHT in children younger than 3.<sup>11</sup> Of the 9 authors we contacted, 2 did not respond, 1 was unable to help, and the authors of the remaining 6 studies, all published since 2003,<sup>9,12–16</sup> generously provided anonymized, individual-level patient information. We have used these data to perform an individual-patient data pooled-analysis to relate combinations of clinical features to AHT or nAHT.

All studies included children younger than 3 who were admitted to hospital with an ICI, defined as any combination of subdural hemorrhage, subarachnoid hemorrhage, extradural hemorrhage, intraparenchymal injury, cerebral contusion, diffuse axonal injury, hypoxic ischemic injury and/or associated cerebral edema. The inclusion criteria with respect to ICI type varied between studies; cases of isolated skull fracture with no ICI were excluded. In 2 instances,<sup>15,16</sup> the authors provided us with additional data over and above that which they had published in the referenced studies. All extra cases had been ascertained in the same manner as that described in the published study.

We defined the different features as follows: RHs included any pattern of RH that was clearly documented on ophthalmological examination. We confirmed an absence of RH when the eyes had been examined and no hemorrhages found; otherwise the data were deemed “missing.” “Long-bone fractures” included metaphyseal or diaphyseal fractures, both old and new, affecting any of the long bones. We classified a fracture as “absent” when there were no fractures reported on a skeletal survey; if no skeletal survey was performed, fracture data were re-

**TABLE 1** Abuse Ranking

Ranking	Criteria Used to Define Abuse
1	Abuse confirmed at case conference or civil, family, or criminal court proceedings or admitted by perpetrator or independently witnessed
2	Abuse confirmed by stated criteria, including multidisciplinary assessment
3	Diagnosis of abuse defined by stated criteria
4	Abuse stated as occurring, but no supporting detail given as to how it was determined
5	Abuse stated simply as "suspected"; no details on whether it was confirmed

corded as "missing." "Seizure" included documented seizures, either single or, in some cases, status epilepticus. "Apnea" and "head/neck bruising" were simply recorded as "present," "absent," or "missing."

We defined AHT (terminology as recommended in the recent American Academy of Pediatrics' recommendations<sup>17</sup>) as ICI where abuse had been confirmed as a cause. We only included confirmed cases of AHT, ie, those ranked 1 or 2 for abuse, according to our previously published 'rank of abuse' (Table 1). This was to ensure that the decision was based on a thorough child protection assessment and multidisciplinary decision or court proceedings to minimize circularity by dependence purely on the clinical features alone to determine if abuse had taken place.<sup>11</sup> All the nonabused children had confirmed causes (traumatic or organic); any cases deemed indeterminate or simply "suspected abuse" in the original data set were excluded from the analysis.

For full details of the statistical methods employed, see the Appendix. In summary, we performed an aggregate analysis of the 6 studies, taking into account the uncertainty arising from missing information on certain clinical features. To do so, we employed multi-

ple imputation, using a bespoke hot-deck imputation strategy informed by examination of the raw data. Wherever possible, we imputed data from another child in the same or a similar study, particularly when several items were missing. If fewer than 5 children could be found to match the child who had missing information, we refrained from imputing from the matched cases because this can lead to underestimation of the uncertainty surrounding the missing features. Instead, we completed our imputed data by filling in any remaining missing items from the margins of the observed distribution of that feature, within their etiology group. We generated 10 imputed data sets for analysis. To each of the imputed data sets, we fitted a multilevel logistic regression model. Logistic regression focuses on the probability of AHT, given information on the presence or absence of certain features. A multilevel version was used to account for the fact that the different studies were drawn from different populations and, in particular, had different prevalences of AHT. Results from the 10 imputed data sets were combined according to established procedures, and expressed as ORs, with 95% confidence intervals (CIs) and PPVs. We emphasize that a large OR need not correspond to a large PPV because a PPV takes into account the underlying prevalence of AHT in this specific population.

We fitted 3 multilevel logistic regression models. Initially, we examined the relationship between AHT, age, and gender, adjusting for variability in the prevalence of AHT between studies. We then added all 7 clinical features to the model, further adjusting for between-study variation in the dependence of etiology (AHT or nAHT) on each clinical feature. To improve statistical efficiency, we then refitted this model, omitting nonsignificant features. Fi-

nally, we assessed the predictive accuracy of the model. To do this, we used fivefold cross-validation: this process divides the data into 5 parts, using 4 to refit the multilevel model, and the remaining part to test its predictive accuracy. This process is then repeated, with each of the 5 parts being used to test the model in turn.

## RESULTS

The 6 included studies<sup>9,12–16</sup> represent data on 1053 children (348 sustained AHT, 705 sustained nAHT). The study designs were similar; they were all population-based and included children younger than 3 who were admitted to regional hospitals with ICI diagnosed on neuroimaging. Four studies included all children with traumatic brain injury,<sup>12–14,16</sup> and 2 included all children who were admitted with subdural hemorrhage.<sup>9,15</sup> Studies confirmed etiology of nAHT cases but applied different exclusion criteria. Two of these studies were prospective,<sup>12,16</sup> with the remainder retrospective<sup>9,13–15</sup> but with detailed consecutive case ascertainment.

### Analysis

Because demographic information was available for all children, it was possible to examine the diagnostic utility of age and gender without resorting to multiple imputation. No significant gender differences were observed, although it was clear that more boys than girls sustained brain injury, regardless of etiology (Table 2). All 3 of the youngest age groups (0–5 months, 6–11 months, 12–23 months) had higher prevalence of AHT than the eldest group (24–36 months), and the differences between the age groups were significant (Table 3). The odds in the 2 youngest groups were ~4 times the baseline odds. Consistent with study ascertainment, we also observed substantial between-study variability in the prevalence of abuse.

**TABLE 2** Univariate Summaries According to Study

Features	Bechtel et al <sup>12</sup>		Ettaro et al <sup>13</sup>		Hettler and Greenes <sup>14</sup>		Hobbs et al <sup>9</sup>		Kemp et al <sup>15</sup>		Vinchon et al <sup>16</sup>	
	n	%	n	%	n	%	n	%	n	%	n	%
Head trauma												
nAHT	50	78	316	84	216	69	31	48	67	40	25	36
AHT	14	22	61	16	95	31	34	52	100	60	44	64
Age group, mo												
0–5	31	48	121	32	202	65	34	52	51	31	45	65
6–11	20	31	77	20	68	22	18	28	33	20	14	20
12–23	10	16	79	21	41	13	4	6	53	32	10	14
24–35	3	5	100	27	0	0	9	14	30	18	0	0
Gender												
Female	31	48	155	41	106	34	29	45	58	35	23	33
Male	33	52	222	59	205	66	36	55	109	65	46	67
Skull fracture												
Absent	28	44	154	41	149	48	0	0	142	85	57	83
Present	36	56	220	58	162	52	0	0	25	15	12	17
Data missing	0	0	3	1	0	0	65	100	0	0	0	0
Rib fracture												
Absent	16	25	358	95	306	98	0	0	141	84	52	75
Present	3	5	14	4	5	2	0	0	26	16	17	25
Data missing	45	70	5	1	0	0	65	100	0	0	0	0
Long-bone fracture												
Absent	17	27	343	91	294	95	0	0	138	83	57	83
Present	2	3	29	8	17	5	0	0	29	17	12	17
Data missing	45	70	5	1	0	0	65	100	0	0	0	0
RH												
Absent	45	70	328	87	225	72	35	54	96	57	33	48
Present	16	25	44	12	86	28	30	46	56	34	36	52
Data missing	3	5	5	1	0	0	0	0	15	9	0	0
Head and/or neck bruising												
Absent	30	47	358	95	0	0	0	0	107	64	35	51
Present	28	44	19	5	0	0	0	0	60	36	34	49
Data missing	6	9	0	0	311	100	65	100	0	0	0	0
Apnea												
Absent	0	0	0	0	0	0	57	88	116	69	51	74
Present	0	0	0	0	0	0	8	12	36	22	18	26
Data missing	64	100	377	100	311	100	0	0	15	9	0	0
Seizures												
Absent	46	72	0	0	235	76	46	71	84	50	33	48
Present	8	12	0	0	76	24	19	29	66	40	36	52
Data missing	10	16	377	100	0	0	0	0	17	10	0	0

**TABLE 3** Regression Analysis of Demographic Information: Global Test of Significance of Age

	Regression Coefficient	Lower Confidence Limit	Upper Confidence Limit	OR	Lower Confidence Limit	Upper Confidence Limit	P	Random Effects SD
Intercept	-1.943	-3.050	-0.835	0.143	0.047	0.434	.001	1.135
0–5 mo	1.500	0.428	2.572	4.480	1.534	13.088	<.001	1.074
6–11 mo	1.399	0.731	2.066	4.051	2.078	7.897	—	<0.001
12–23 mo	0.752	-0.067	1.571	2.121	0.935	4.813	—	0.483
Male	0.280	-0.021	0.580	1.322	0.979	1.787	.069	<0.001

### Multiply Imputed Multilevel Logistic Regression

For each of the 10 imputed data sets, we used multilevel logistic regression to determine the diagnostic value of all

features identified during initial investigations in combination. We report effect sizes on the logistic scale and in terms of ORs, giving 95% CIs for these estimates, and PPVs, together with *P*

values and estimates of the SDs of the random effects (Table 4). In this model, the reference category is a girl aged between 24 and 36 months, but now, additionally, they are known to have no other clinical features present (ie, ICI alone).

### Nonsignificant Features

Once other clinical features were known, age was no longer significantly associated with etiology at the 5% level. In other words, given information about the presence or absence of the 7 clinical features, knowing the child's

**TABLE 4** Full Regression Model of Demographic and Clinical Features Expressed as ORs With 95% CIs, *P* Values, and Estimates of the SDs of the Random Effects

	Regression Coefficient	Lower Confidence Limit	Upper Confidence Limit	OR	Lower Confidence Limit	Upper Confidence Limit	<i>P</i>	Random Effects SD
Intercept	-4.002	-5.505	-2.500	0.018	0.004	0.082	<0.001	1.238
0–5 mo	1.138	-0.339	2.615	3.120	0.712	13.667	0.068	1.371
6–11 mo	1.284	0.277	2.291	3.612	1.320	9.884	—	<0.001
12–23 mo	0.267	-1.093	1.627	1.306	0.335	5.088	—	0.910
Male	0.149	-0.435	0.732	1.160	0.647	2.079	0.616	0.194
Skull fractures	-0.749	-1.928	0.430	0.473	0.145	1.537	0.213	1.142
Rib fractures	3.755	1.963	5.548	42.746	7.119	256.668	<0.001	0.446
Long-bone fractures	2.635	1.179	4.092	13.949	3.250	59.865	<0.001	0.971
RHs	3.450	2.824	4.076	31.497	16.840	58.912	<0.001	0.159
Head and/or neck bruising	1.513	0.160	2.865	4.539	1.174	17.546	0.027	1.115
Apnea	1.938	0.742	3.135	6.948	2.100	22.985	0.001	0.617
Seizures	1.624	0.675	2.574	5.075	1.963	13.117	0.001	0.762

age as well did not significantly alter the probability that the child had been abused. Gender remained uninformative where AHT was concerned. Skull fracture(s) were the only clinical feature(s) with an OR of <1 (ie, more suggestive of nAHT than AHT), although this association did not reach significance.

Because we found no statistical evidence that including these terms in a model for AHT was important, we refitted the model, dropping the nonsignificant features. The reduced, multiply-imputed multilevel analysis gave rise to Table 5, on which we base our substantive conclusions.

### ICI Alone

When a child younger than 3 had an ICI, with none of the clinical features noted above, the estimated probability of AHT

was ~4% (Table 5). Although we did account for study-specific associations between AHT and age, AHT and gender, and AHT and skull fractures, we reiterate that age did not help to determine the likelihood of abuse when all other clinical features were known.

### Results When Each Feature Is Solely Present in a Child With ICI

#### Fractures

For rib fractures there was the strongest evidence of AHT, with an OR of ~45 (Figs 1 and 2). Note, however, the wide CIs that surround this estimate, owing to the very small numbers of rib fractures present in both etiology groups. Long-bone fractures are similarly indicative of AHT (OR: 13.75), slightly less strongly than rib fractures. Because

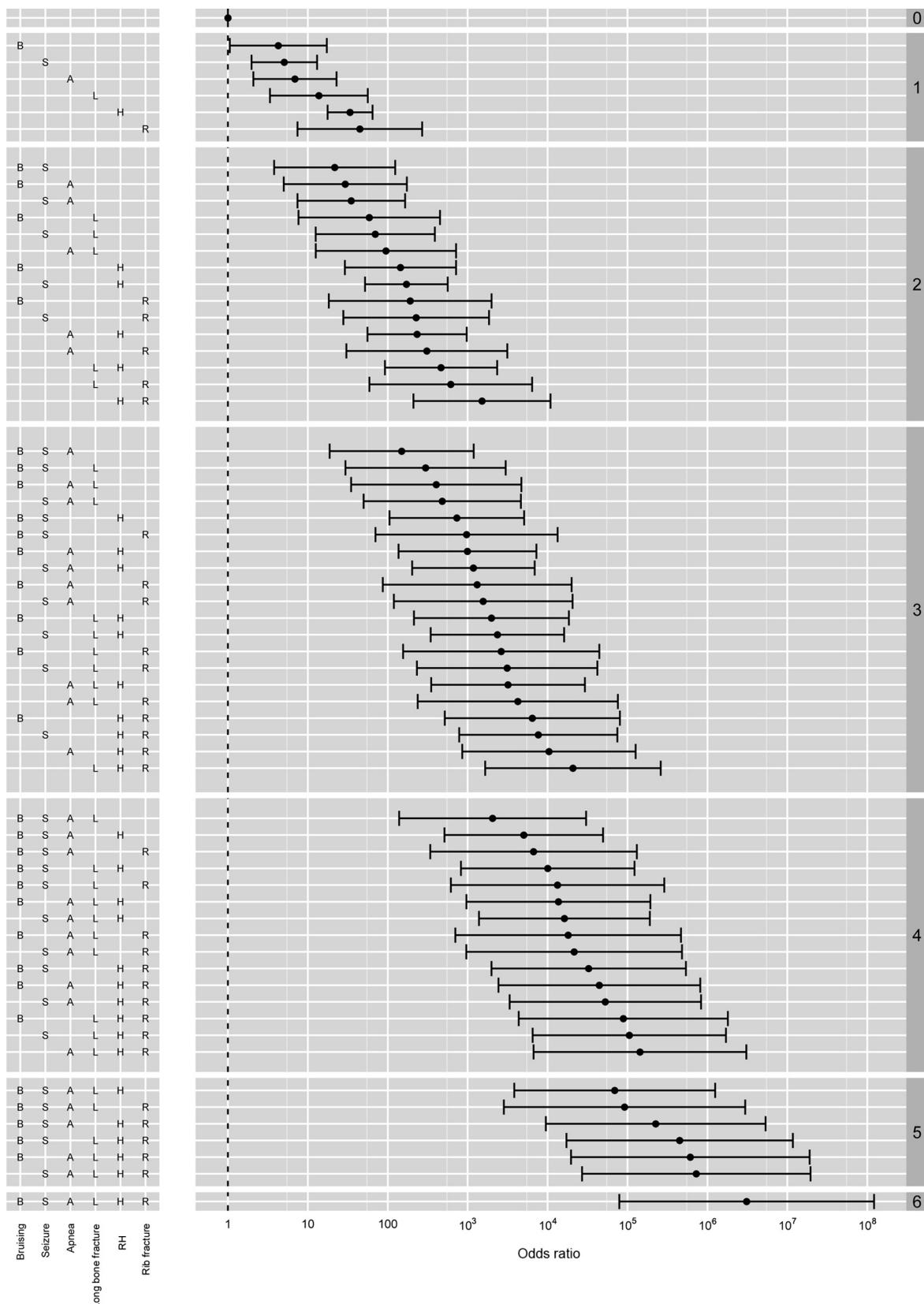
the prevalence of long-bone fractures was marginally higher than that of rib fractures, the CIs surrounding this estimate are narrower.

#### *RH, Head and/or Neck Bruising, Apnea, and Seizures*

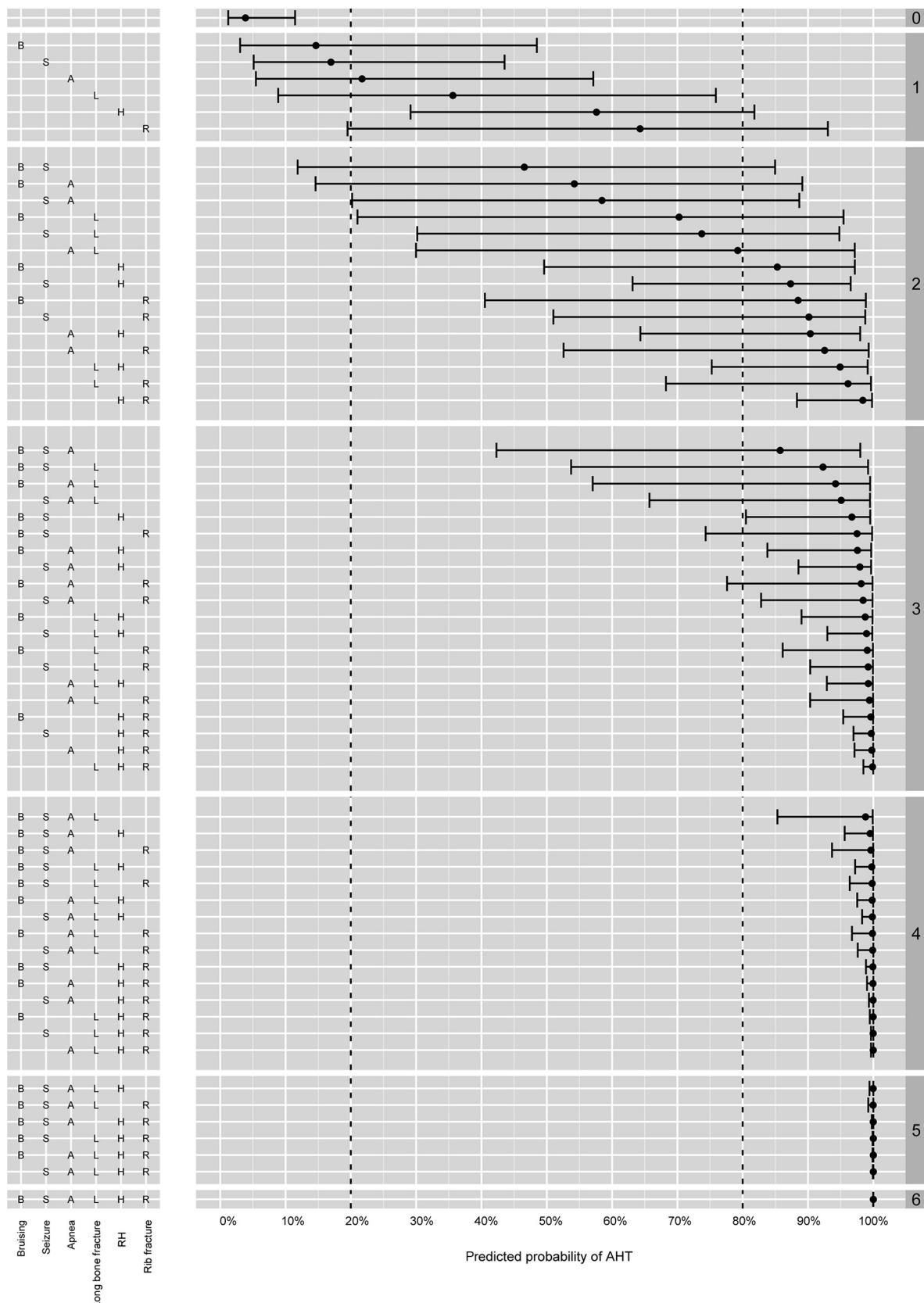
RHs were 1 of the more frequently recorded items, and had an estimated OR of ~35. Therefore, whereas a child with an ICI alone had a probability of AHT of 4%, this rose to 58% if ICI and only RHs were present. Head and/or neck bruising had a weaker, marginally significant relationship with AHT. A combination of ICI and bruising to the head and/or neck increased the probability of AHT from 4% to 15%. It is worth emphasizing, however, that this left 85% of children with this pattern of injury who did not have AHT.

**TABLE 5** Reduced Regression Model

	Regression Coefficient	Lower Confidence Limit	Upper Confidence Limit	OR	Lower Confidence Limit	Upper Confidence Limit	<i>P</i>	Random Effects SD
Intercept	-3.213	-4.358	-2.067	0.040	0.013	0.127	<0.001	1.117
0–5 mo								1.210
6–11 mo								0.814
12–23 mo								1.069
Male								0.160
Skull fracture								1.240
Rib fractures	3.800	2.036	5.565	44.720	7.659	261.098	<0.001	0.311
Long-bone fractures	2.621	1.235	4.006	13.747	3.440	54.937	<0.001	0.930
RHs	3.519	2.882	4.157	33.764	17.855	63.848	<0.001	0.211
Head and/or neck bruising	1.451	0.072	2.830	4.268	1.075	16.951	0.038	1.175
Apnea	1.931	0.732	3.129	6.893	2.079	22.858	0.001	0.623
Seizures	1.624	0.695	2.553	5.072	2.003	12.843	0.001	0.737

**FIGURE 1**

OR for AHT in children younger than 3 years who have an ICI (log scale).



**FIGURE 2**

PPV of AHT in children younger than 3 years who have an ICI.

Either apnea or seizures alone, in association with ICI, significantly increased the likelihood of AHT, both at a level comparable to that of head and/or neck bruising alone (an OR of 5). Apnea was the slightly stronger of the 2, although this finding is subject to a degree of uncertainty because of the large amount of missing data (recorded in only 301 of 1053 cases). It is hoped that this feature will be more commonly recorded in future studies because of increasing recognition of its association with AHT.

### **Results When Multiple Features Are Present**

A unique advantage of our study is the ability to consider the presence or absence of multiple clinical features in combination. Thus far, we have used the data to focus on situations where we knew that only a single clinical feature was present; however, we now turn to the more frequent scenario where more than 1 feature may be present (Figs 1 and 2). It is immediately apparent that when only 1 or 2 features were present, the ability to distinguish AHT from nAHT depends heavily on the specific feature(s) in question. For instance, if a child with ICI had head and/or neck bruising in combination with apnea, the estimated probability of AHT was 54% (OR: 29); however, if the child had apnea combined with RHs, this estimated probability rose to 90%, and the OR was almost 10 times higher. Likewise, if the child presented with apnea and seizures but no other features, the estimated probability of AHT was only 58% (OR: 35); however, if a child had seizures and rib fractures, the probability of AHT rose to 90%, and the OR was 227.

Once 3 or more of the significant features were present, ORs were  $>100$ , and PPVs for AHT were uniformly above 85%, irrespective of the specific features. In Fig 2 we show precise esti-

mates for each of the 64 possible combinations.

### **Accuracy of the Model: Model Checking**

In the absence of an additional, independent source of data against which to test our findings, we used cross-validation to assess the accuracy of our predictions. Details of the model checking are provided in the Appendix. We chose high cut-off limits, such that we deemed a predicted probability to be correct if the PPV for AHT or nAHT was  $>80\%$ , and the predicted etiology was true (see Fig 2). Any PPV less than this was regarded as indeterminate. On this basis, our predicted etiology was correct 80% of the time, indeterminate 15% of the time, and incorrect in 5% of cases. This highlights that no set of clinical features was unique to AHT or nAHT, and such features must be considered in the context of all other medical and social aspects of the case in question.

### **DISCUSSION**

Our analysis of more than 1000 infants and young children with ICI of confirmed etiology represents the largest published analysis of combined clinical features to estimate the probability of AHT. We have shown that in a child younger than 3 with an ICI and 1 or 2 of the key clinical features, the probability of AHT varied depending on the number and specific features present, with RHs and rib fractures being the most discriminating. Three or more of the key features were highly predictive of AHT. This analysis offers the potential to underpin a clinical opinion with a valid scientific estimate of probability.

In the study we draw on the raw data from the most recent large-scale comparative epidemiologic publications, and we are indebted to the authors of the primary studies included. This is the first detailed multivariate analysis

to be produced and, in common with all studies in this difficult field, there are limitations; however, we believe that this work makes a timely and valuable contribution to the clinical field, in particular at a time when the validity of diagnosing AHT from combinations of clinical features is being questioned.<sup>3</sup> Although information on the presence of apnea was missing in a large number of cases, the majority of these (377 + 311) came from 2 large studies that had otherwise very complete recording of features. Where data (eg, apnea) was missing, we addressed this by using similar individuals within the data set to impute missing data, an approach that is statistically valid provided the data are missing at random.<sup>18</sup> In statistical terms, clinical decisions about what investigations to perform define the mechanism leading to missing data. Because such decisions are usually taken sequentially and on the basis of the results of the previous investigations (that is, the observed data), the "missing at random" assumption is reasonable in this context. All of the features, other than apnea, were investigated for in the majority of the studies.

The difference between the included studies highlights the varied approach to the clinical assessment of these children. The overall study findings reiterate the importance of a comprehensive and standardized investigation of all young children with an ICI of uncertain etiology, recording the presence or absence of seizures, apneic episodes, bruising, and findings from fundoscopy and skeletal survey to derive an initial clinical probability of AHT. The cases included in this analysis were investigated in the late 1990s and early 2000s. Undoubtedly, future studies are likely to include magnetic resonance imaging, more rigorous skeletal survey protocols and full ophthalmological data, each of which potentially

reveals a greater understanding of the individual diagnostic indicators of AHT. However, the features included in our analysis are those that all frontline clinicians are likely to identify before referring the child into a specialist center for more detailed assessment, and, as such, the results of this study could be used to support such additional detailed assessments.

This model has been developed from recent high quality, large scale comparative studies currently available from the international scientific literature. However, we believe that it has the potential for future development, into a sophisticated tool to aid the clinical decision process when assessing a young child with unexplained head trauma. A large-scale prospective study is urgently needed to collect a more extensive standardized clinical data set from a cross-section of young children with ICI and apply a multivariate analysis. A proposed study of this nature would need to be conducted on an international basis to collect sufficient case numbers over a reasonable time scale and to incorporate more detail about the nature of the currently recognized clinical indicators such as seizure type and duration, the precise location and pattern of RHs,<sup>19,20</sup> and detailed neuroradiological features of ICI found in each group.<sup>21</sup> Other features may also be contributory eg, history on presentation,<sup>14</sup> vomiting,<sup>22</sup> co-existent injury,<sup>1,23</sup> conscious level on arrival,<sup>16</sup> and burns<sup>1</sup>; with increasing features included in the analysis, larger numbers will be required specifically to ensure that enough cases with each possible combination (features present or absent) are included to enable valid statistical analysis of every scenario.

Diagnostic studies in this field are open to criticism of circularity because of their dependence on a constellation of clinical features, as op-

posed to a single gold-standard diagnostic test, which does not exist. Ultimately, in any individual case, a child either has, or has not, suffered AHT and, consequently, a diagnosis of AHT either is, or is not, correct. However, except in cases of independently witnessed injury, a diagnosis must rest on a probabilistic assessment of how likely it is that AHT took place. It is not possible to restrict research in this field to independently witnessed abuse, which represents a tiny proportion of cases. We have attempted to minimize the risk of circularity by only analyzing cases where abuse was confirmed by a comprehensive evaluation of all the medical and social features, after a multidisciplinary assessment of the full case details and, in many cases, by "finding of fact" in care or criminal legal proceedings or perpetrator admissions. Likewise, nAHT cases were only included when the etiology was confirmed. To adequately power studies of what is, in statistical terms, a small population, some authors frequently combine presumptive and 'suspected' abuse into 1 category.<sup>13</sup> Others<sup>10</sup> add indeterminate cases to the noninflicted cases or combine suspected abuse with confirmed abuse. We have rigidly excluded such cases from our analysis and the extremely large data set and internal cross validation offers reassurance that our estimates are valid. In addition, the strength of this work lies in a regression analysis of 6 significant variables in 1053 children, generating 64 possible combinations, which represents a wider range of possible combinations than could be achieved from any single study. The findings from this comparative study mitigate against the circularity argument to some extent. We have demonstrated that there is a difference between the predictive probability of different combinations of features, strongly influenced by the specific features in question. However,

even for strongly influential features, such as RHs, not all cases were because of AHT, as shown by a PPV of 58% for those children with an ICI and RHs and no other clinical features. These data includes the most challenging clinical scenario, namely the likelihood of abuse when a child has ICI but none of the other distinguishing clinical features. For a child younger than 3 with an ICI alone, the estimated probability that the brain injury is an AHT is 4% in this data set.

Although the use of this analysis to determine the probability of AHT in a given case will never replace the diagnostic skills of the clinician, it has the potential to contribute to decision-making. This could assist frontline professionals when deciding whether to refer a child for specialist clinical and multiagency investigation of possible AHT and contribute to decision-making at various points along the referral and assessment process. It could also assist clinicians offering medical testimony in civil or criminal proceedings, in demonstrating why certain combinations of features are more or less predictive of an abusive etiology. Contrary to the view expressed by Tuerkheimer<sup>3</sup> that "the scientific underpinnings of SBS have crumbled over the past decade," this large-scale analysis confirms the association of AHT with specific combinations of clinical features and, furthermore, it has enormous potential as a prototype for a larger-scale prospective study to include more sophisticated details of clinical indicators, which would be valuable to clinicians and other professionals working in this field.

## ACKNOWLEDGMENTS

This work was supported by the National Society for the Prevention of Cruelty to Children, the Royal College of Pediatrics and Child Health, and Wales Office of Research and Development.

Data were provided by Dr Kemp; K. Bechtel, Yale University (New Haven, CT); L. Et-taro, University of Pittsburgh (Pitts-

burgh, PA); J. Hettler, Baystate Medical Center (Springfield, MA); C. Hobbs, St James's University Hospital (Leeds, West

Yorkshire, United Kingdom); and M. Vinchon, Centre Hospitalier Régional Universitaire (Lille, France).

## REFERENCES

- 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.
- Oral R, Yağmur F, Nashelsky M, Turkmen M, Kirby P. Fatal abusive head trauma cases: consequence of medical staff missing milder forms of physical abuse. *Pediatr Emerg Care*. 2008;24(12):816–821.
- Tuerkheimer D. The Next Innocence Project: shaken baby syndrome and the criminal courts. *Wash Univ Law Rev*. 2009;87(1):1–58. Available at: <http://lawreview.wustl.edu/inprint/87/1/dtuerkheimer.pdf>. Accessed April 1, 2011.
- Tuerkheimer D. Anatomy of a misdiagnosis. *The New York Times*. September 20, 2010. Available at: [www.nytimes.com/2010/09/21/opinion/21tuerkheimer.html](http://www.nytimes.com/2010/09/21/opinion/21tuerkheimer.html). Accessed April 1, 2011.
- Barr M, Leventhal JM. A spotlight on shaken baby syndrome. *The New York Times*. September 24, 2010. Available at: [www.nytimes.com/2010/09/25/opinion/25babyl.html](http://www.nytimes.com/2010/09/25/opinion/25babyl.html). Accessed April 1, 2011.
- Kennedy H. Sudden unexpected death in infancy: a multi-agency protocol for care and investigation. The Royal College of Pathologists and The Royal College of Paediatrics and Child Health; September 2004. Available at: [www.rcpath.org/index.asp?pagelD=455](http://www.rcpath.org/index.asp?pagelD=455)
- Ewing-Cobbs L, Kramer L, Prasad M, et al. Neuroimaging, physical, and developmental findings after inflicted and noninflicted traumatic brain injury in young children. *Pediatrics*. 1998;102(2 pt 1):300–307.
- Wells RG, Vetter C, Laud P. Intracranial hemorrhage in children younger than 3 years: prediction of intent. *Arch Pediatr Adolesc Med*. 2002;156(3):252–257.
- Hobbs C, Childs AM, Wynne J, Livingston J, Seal A. Subdural haematoma and effusion in infancy: an epidemiological study. *Arch Dis Child*. 2005;90(9):952–955.
- Hoskote A, Richards P, Anslow P, McShane T. Subdural haematoma and non-accidental head injury in children. *Childs Nerv Syst*. 2002;18(6–7):311–317.
- Maguire S, Pickerd N, Farewell D, Mann M, Tempest V, Kemp AM. Which clinical features distinguish inflicted from non-inflicted brain injury? A systematic review. *Arch Dis Child*. 2009;94(11):860–867.
- 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.
- Ettaro L, Berger RP, Songer T. Abusive head trauma in young children: characteristics and medical charges in a hospitalized population. *Child Abuse Negl*. 2004;28(10):1099–111.
- Hettler J, Greenes DS. Can the initial history predict whether a child with a head injury has been abused? *Pediatrics*. 2003;111(3):602–607.
- Kemp AM, Stoodley N, Cobley C, Coles L, Kemp KW. Apnoea and brain swelling in non-accidental head injury. *Arch Dis Child*. 2003;88(6):472–476.
- Vinchon M, Defoort-Dhellemmes S, Desmont M, Dhellellemmes P. Accidental and non-accidental head injuries in infants: a prospective study. *J Neurosurg*. 2005;102(suppl 4):380–384.
- Christian CW, Block R, Committee on Child Abuse and Neglect, American Academy of Pediatrics. Abusive head trauma in infants and children. *Pediatrics*. 2009;123(5):1409–1411.
- Little RJA, Rubin DB. *Statistical Analysis With Missing Data*. 2nd ed. Hoboken, NJ: John Wiley & Sons Inc; 2002.
- Duharre AC, Christian C, Armonda R, Hunter J, Hertle R. Disappearing subdural hematomas in children. *Pediatr Neurosurg*. 1996;25(3):116–122.
- Pierre-Kahn V, Roche O, Dureau P, et al. Ophthalmologic findings in suspected child abuse victims with subdural hematomas. *Ophthalmology*. 2003;110(9):1718–1723.
- Morad Y, Kim YM, Mian M, Huyer D, Capra L, Levin AV. Nonophthalmologist accuracy in diagnosing retinal hemorrhages in the shaken baby syndrome. *J Pediatr*. 2003;142(4):431–434.
- Tzioumi D, Oates RK. Subdural hematomas in children under 2 years: accidental or inflicted? A 10-year experience. *Child Abuse Negl*. 1998;22(11):1105–1112.
- Shugerman RP, Paez A, Grossman DC, Feldman KW, Grady MS. Epidural hemorrhage: is it abuse? *Pediatrics*. 1996;97(5):664–668.

## APPENDIX: DETAILS OF STATISTICAL METHODS

We began our analysis by visualizing the raw data with a parallel-coordinates plot (Fig A1). Each child is represented in this plot by a connected line that joins the recorded values of the variables of interest. Variables are transformed to share a common scale, and their values can be read off the vertical axes. The plot is then divided by injury etiology and by the source study, and lines are jittered and colored to alleviate problems with overplotting. Densely colored areas of the plots represent more children than the light, sparse areas. For example, the predominant pattern of features seen in the AHT group of children in the Vinchon et al<sup>1</sup> study is the youngest age group, boys, without skull, rib, or long-bone fractures, with retinal haemorrhages, no information on bruising and apnea but with seizures.

Inspection of the raw data informed our hotdeck multiple-imputation strategy. The Kemp et al<sup>2</sup> study had complete information on all children and, thus, required no imputation. Hobbs et al<sup>3</sup> had only a small amount of missing data, and (where at least 5 matches could be found) this was imputed from either itself or from the Kemp et al study.<sup>2</sup> Ettaro et al<sup>4</sup> had a small amount of missing fracture data, and neither apnea nor seizures were recorded for these children. The fracture data we imputed from within the same study matched on all available features. This being done, we then found children in the Vinchon et al study<sup>1</sup> who matched the Ettaro et al sample<sup>4</sup> on all features common to both studies, using this to impute information on seizures. A similar strategy was used to bring information on bruising from the Ettaro et al study<sup>4</sup> into the Vinchon et al<sup>1</sup> cases. At this stage, both Ettaro et al<sup>4</sup> and Vinchon et al<sup>1</sup> had complete information except for apnea (always excepting

any instances in which insufficient matching children could be found).

After a within-Bechtel et al<sup>5</sup> imputation to fill in the sporadically missing items, data from (the imputed versions of) the Ettaro et al<sup>4</sup> and Vinchon et al<sup>1</sup> studies were used to update the Bechtel et al<sup>5</sup> data. At this stage Bechtel et al,<sup>5</sup> Ettaro et al,<sup>4</sup> and Vinchon et al<sup>1</sup> were all used to form a sampling frame from which to impute into Hettler and Greenes,<sup>6</sup> who chose not to record many items but, importantly, recorded information on the presence or absence of apnea. At this point Hettler and Greenes,<sup>6</sup> Hobbs et al,<sup>3</sup> and Kemp et al<sup>2</sup> had (in principle) complete information, whereas Bechtel et al,<sup>5</sup> Ettaro et al,<sup>4</sup> and Vinchon et al<sup>1</sup> lacked data only on apnea. By sampling information on apnea from similar children in the former set into the latter, we completed the hotdeck stage of imputation.

In practice, there were several instances in this process where fewer than 5 matching children could be found. The choice of a minimum of 5 matched individuals is somewhat arbitrary, but it is reasonably likely that 4 or fewer children may have exactly the same observed information on the missing item(s), especially if the prevalence of the relevant feature is particularly low or particularly high. Thus, imputing from such a sample could erroneously suggest that the missing item(s) were, in fact, known. To circumvent this problem, the final stage of our imputation was to sample from the margin of the feature concerned conditional on the etiology of the neurologic injury.

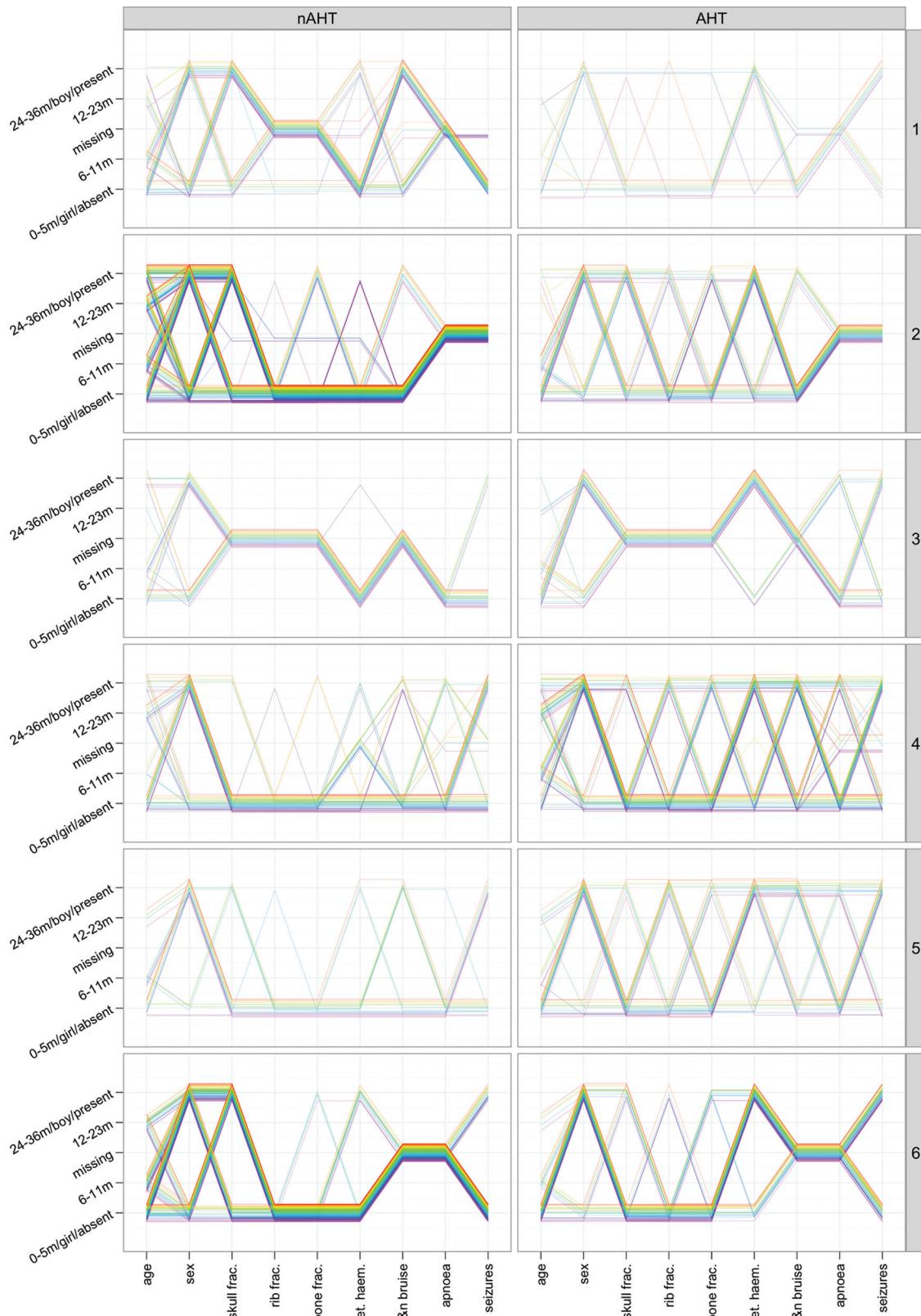
Figure A2 displays the 10 imputed data sets, made slightly transparent so that similarities and differences between imputations may be seen. Note that results of the 2 subdural-only studies (Hobbs et al<sup>3</sup> and Kemp et al<sup>2</sup>) do not show the propensity for skull fractures in the nAHT group seen (or im-

puted) in the other 4 studies. This may be because of a number of organic causes of subdural hemorrhage in this group, whereas the nAHT groups in the other studies were trauma-related. An interesting feature of the Hettler and Greenes<sup>6</sup> imputations is that rib and long-bone fractures are never imputed to be present, in contrast with skull fractures and bruising. Another obvious difference between the 2 etiology groups is the number of isolated features in the nAHT groups as opposed to the high prevalence of multiple recorded features in the AHT groups.

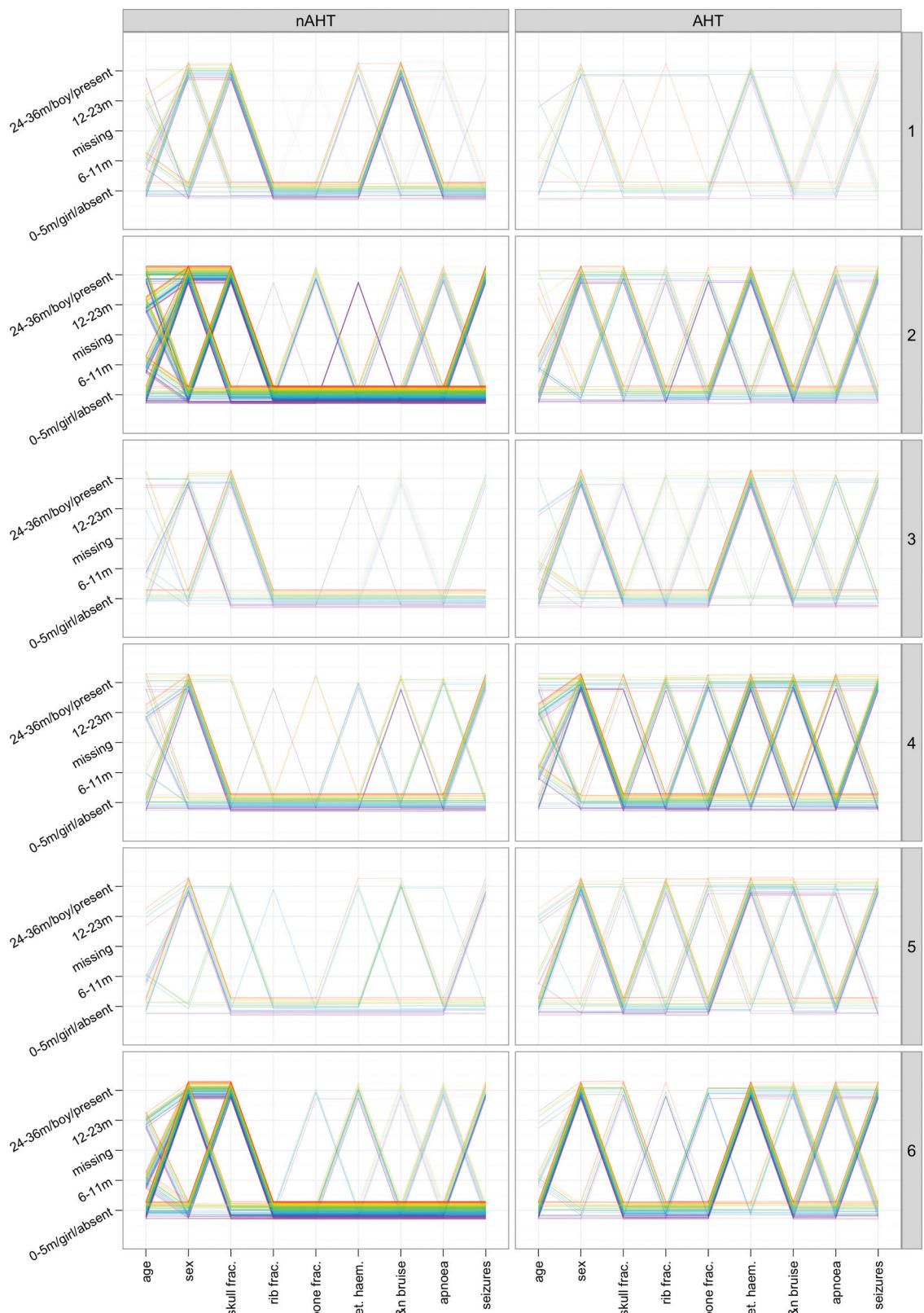
Following the modelling described in the article, we gave detailed consideration to the predictive performance of our chosen model. Figure A3 shows the raw data again but this time colored according to the predicted probability of AHT (based on the final, reduced model). Because there was substantial missing information, these predictions were averaged across the 10 imputed data sets. A green color suggests AHT with a probability less than 20%; orange denotes ambiguous cases with predicted probabilities of AHT that are between 20% and 80%; and a red color corresponds to high probabilities (>80%) of AHT.

We see that our modelling is reasonably successful at discriminating between the 2 groups of interest. The predominance of green and orange in the left-hand plots and the frequency of red lines in the right-hand plots suggest that, on the whole, the model is fitting quite well. However, of (at least) equal importance is the degree to which the model can suggest incorrect etiologies; it is possible to have a rib fracture in the nAHT group; conversely, some children with no obvious injury (particularly evident in the Hobbs et al<sup>3</sup> study) can have AHT.

As mentioned in "Methods," we recognize the risk of overestimating the risk of abuse by relying solely on within-

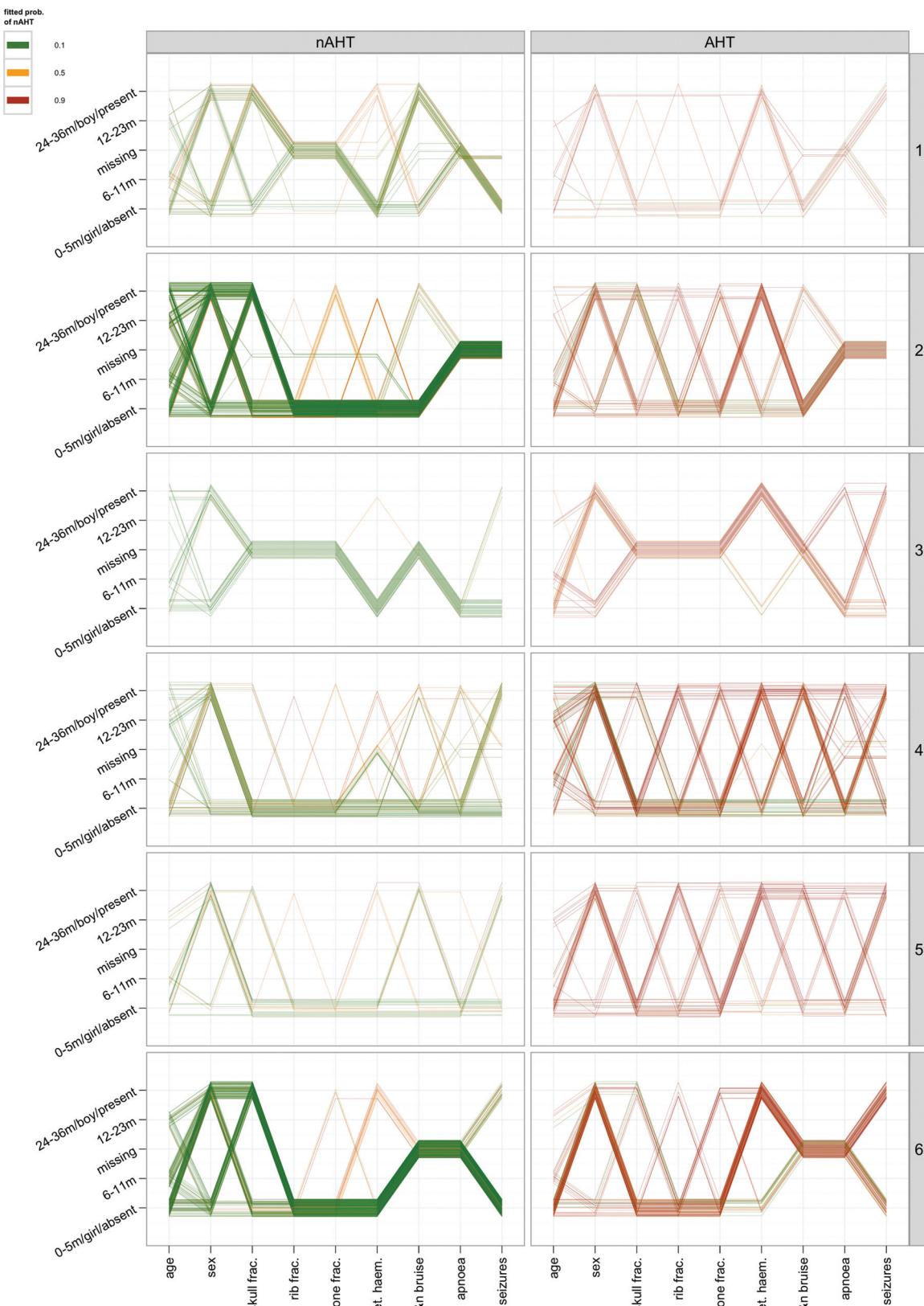
**FIGURE A1**

Parallel-coordinates plot of the raw data, in which each child's information is represented by a connected line.

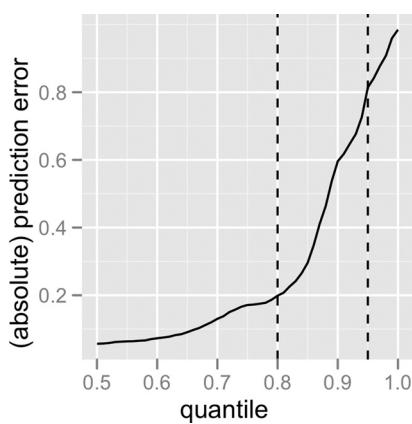


**FIGURE A2**

Parallel-coordinates plot of the 10 imputed data sets.

**FIGURE A3**

Parallel-coordinates plot showing also the probability of abuse for each child in the data set, estimated using the final logistic regression model. A red line denotes an estimated probability of abuse of at least 80%, a green line denotes a probability of abuse of at most 20%, and an orange line is shown in all other cases.



**FIGURE A4**

Quantiles of absolute prediction error. The reference lines show that 80% of the data has a prediction error of  $<0.2$ , whereas 5% of the data has a prediction error of  $>0.8$ .

case data. To address this concern, we conducted a cross-validation investigation to attempt to quantify our ability to predict AHT. We used fivefold cross-validation, whereby the data is parti-

tioned (at random) into 5 subsets. Each of these 5 subsets is used in turn as the test data, and the reduced model is refitted by using the other 4 subsets as the training data.

Results were consistent across the 5 cross-validations. Graphing the quantiles of absolute prediction error (Fig A4), we found that in the best 80% of cases the predictions are confident (positive predictive values either  $>80\%$  or  $<20\%$ ) and correctly so. However, in the worst 5% of the cases, the predictions are confident (using the same definition) but incorrect. Important, however, is that most of these confident-but-wrong predictions incorrectly suggest nAHT, which highlights the need for any assessment of likelihood of abuse to be based on a full multidisciplinary assessment rather than relying on clinical features alone.

- Vinchon M, Defoort-Dhellemmes S, Desurmont M, Dhellemmes P. Accidental and non-accidental head injuries in infants: a prospective study. *J Neurosurg*. 2005;102(4 suppl):380–384
- Kemp AM, Stoodley N, Cobley C, Coles L, Kemp KW. Apnoea and brain swelling in non-accidental head injury. *Arch Dis Child*. 2003;88(6):472–476
- Hobbs C, Childs AM, Wynne J, Livingston J, Seal A. Subdural haematoma and effusion in infancy: an epidemiological study. *Arch Dis Child*. 2005;90(9):952–955
- Ettaro L, Berger RP, Songer T. Abusive head trauma in young children: characteristics and medical charges in a hospitalized population. *Child Abuse Negl*. 2004;28(10):1099–1111
- 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
- Hettler J, Greenes DS. Can the initial history predict whether a child with a head injury has been abused? *Pediatrics*. 2003;111(3):602–607