Study Protocol and Statistical Analysis Plan

Official Title: Implementation Trial of a Validated Clinical Decision Rule for Pediatric Abusive Head Trauma

Identifiers: NCT03162354   Unique Protocol ID: STUDY00005613   Secondary IDs: P50HD089922

Document Date: January 29, 2016

SPECIFIC AIMS: Project 2

Abusive head trauma (AHT) is the leading cause of traumatic death and disability in early childhood, affecting at least 690,000 children annually in the United States.[1-7] Physicians who care for young victims of trauma must make important decisions to launch or forgo child-abuse evaluations. These decisions can be difficult, and the stakes are high. Research has shown that AHT was missed or unrecognized in 30% of head-injured children, more than 25% of children with unrecognized AHT suffered additional inflicted injuries when returned to abusive caregivers, and 10% of children with unrecognized AHT subsequently died or were killed—80% of whom could have been saved by earlier recognition of abuse.[8] Moreover, a flawed decision to launch an abuse evaluation can increase parental stress, expose the child to additional risks, prolong hospital stays, and increase health care costs.[9, 10] Unfortunately, physicians routinely struggle to define a "reasonable suspicion" of abuse[11-18] and have demonstrated biases: Younger children of minority race, from single-parent households, and of low socioeconomic status get evaluated for abuse more often.[8, 11-13, 17, 19-21]

The principal investigator (PI; Hymel) conceptualized, designed, and directed sequential multicenter studies to derive and validate an effective AHT screening tool.[22, 23] This screening tool comes in the form of a clinical decision rule (CDR)—an evidence-based tool developed to guide a discrete decision in the continuum of AHT care.[24] Applied at the time of pediatric intensive care unit (PICU) admission, the CDR recommends abuse evaluations for all "high risk" patients defined as presenting with closed head trauma and any one or more of the following: (1) acute respiratory compromise; (2) bruising of the torso, ear(s), or neck; (3) bilateral or interhemispheric subdural hemorrhage(s) or fluid collection(s); and (4) any skull fracture(s) other than an isolated unilateral non-diastatic linear parietal skull fracture.[23] Our preliminary studies of acutely head-injured children (N = 500) admitted to 18 PICUs demonstrated that application of this CDR significantly increased AHT detection (sensitivity 0.96) and decreased unnecessary abuse evaluations (specificity 0.43).[23, 25] Though promising, the CDR has not yet been fully tested in a formal randomized trial.

The proposed CDR clinical trial will demonstrate the CDR's impact on AHT screening accuracy. We hypothesize that CDR application will increase AHT detection and reduce unnecessary abuse evaluations. We will conduct a stratified cluster randomized trial (SCRT) at eight US PICUs randomly assigned to intervention (n = 4) or control (n = 4) conditions. The SCRT will compare AHT screening accuracy after the deployment of active multifaceted implementation strategies designed to promote CDR acceptance, utilization, and accuracy at the intervention sites. These strategies include provider training with onsite visits, site-specific feedback, tailored booster emails to Providers, CDR badge cards, and a mobile-device "AHT Probability Calculator". A clear demonstration of the CDR's positive clinical impacts and an evaluation of avenues to promote sustainability will inform and accelerate subsequent efforts to disseminate the CDR more broadly. The following specific aims are proposed:

Aim 1. Conduct a SCRT across eight PICUs to assess the CDR's impact on AHT screening accuracy. In comparison to non-intervention control sites, we hypothesize that CDR application as an AHT screening tool at the intervention sites will be associated with higher percentages of high-risk patients evaluated for child abuse and lower percentages of low-risk patients evaluated for child abuse.

Aim 2. Identify site-, provider- and patient-specific factors that predict CDR application in PICU settings. We hypothesize that sites with higher patient volumes will demonstrate higher percentages of high-risk patients evaluated for child abuse and lower percentages of low-risk patients evaluated for child abuse. We also hypothesize that Providers with child-abuse expertise, high exposure to the CDR implementation strategies, and high CDR acceptability will demonstrate higher percentages of high-risk patients evaluated for child abuse and lower percentages of low-risk patients evaluated for child abuse. Although we suspect that CDR application will reduce patient disparities at intervention sites, overall, we predict that patients younger than 6 months or those of minority race will be evaluated for abuse more often than patients older than 6 months or nonminority.

Exploratory Aim 3. Once the SCRT has concluded, conduct a 12-month sustainability trial at intervention sites whereby implementation strategies are systematically omitted one-by-one in order to discern their relative importance to sustainability.

Impact. This simple, safe, inexpensive, reliable, readily accessible, evidence-based, and validated CDR will be the first child-abuse screening tool to undergo a formal clinical trial. Applied accurately and consistently, the CDR will decrease the negative impacts of physicians' inherent biases and practice disparities, significantly increase AHT detection, decrease unnecessary abuse evaluations (and their associated risks), and reduce AHT-associated health care costs. Most importantly, the CDR will save lives—the lives of children who will otherwise suffer additional fatal inflicted injuries if/when their AHT is missed or unrecognized.[8, 25,]

A. SIGNIFICANCE

Among all forms of violence directed against infants and young children, AHT is the leading cause of traumatic death and disability, impacting at least 690,000 U.S. children annually. [1-7] To help confirm or exclude AHT masquerading as accidental head trauma, physicians must decide whether or not to launch a workup for child abuse in their young, acutely head-injured patients. The signs and symptoms of accidental head injuries are non-specific and can be the same as those from abuse, associated external injuries are not universal, and decisions are further complicated by the fact that patients are often unable to describe their traumas because they are so young and have head injuries. In addition, caregiver accounts of their children's head injury events can be absent, minimized, changing, or fabricated, and for physicians, discussing abuse with parents is often uncomfortable, and familial psychosocial risk factors can impair physicians' objectivity. Not surprisingly then, physicians often struggle to define a "reasonable suspicion" of abuse[11-18] and have been shown to exhibit biases regarding child maltreatment based on family racial, socioeconomic, and marital status.[8, 11-13, 19-21] In addition, physicians differ in their medical decisions to evaluate young victims of trauma for abuse, with abuse being diagnosed more frequently in large centers than in small centers for patients with equivalent injuries.[8, 12, 17-21] Finally, the validity of AHT as a medical diagnosis is now openly and aggressively disputed in medical journals, in the lay press, and in courtrooms.[26-33] As a result, many physicians are fearful of making unwarranted AHT diagnoses and being challenged in court for doing so, leading them to avoid working up children for abuse in the first place.

A1. The stakes are high. A doctor's flawed decision to forgo an abuse evaluation in an abused head-injured child puts that child at substantial risk for further abuse and/or death when returned to his or her abusive caregiver(s).[8] In their sentinel 1997 study of "missed" AHT, Jenny, Hymel (the PI of this study), and their co-authors reported that (1) AHT was missed or unrecognized in 3 of 10 abused head-injured children when those children first presented for medical care, (2) more than 1 in 4 children with missed or unrecognized AHT suffered additional inflicted injuries when they were returned to their abusive caregiver(s), (3) almost 1 in 10 children with missed AHT subsequently died or were killed, and (4) 4/5 of these deaths might have been prevented by earlier recognition of their abuse.[8]

Launching an unnecessary abuse evaluation in a child who has suffered an accidental head injury can also have adverse consequences, including increased parental stress, strain on the doctor-parent relationship, exposure of the child to additional risks (e.g., radiation, sedation), false positive results that could trigger a misguided child-protection and/or police investigation, prolonged hospital stays, and increased health care costs.[9, 10] Years after Jenny, Hymel, and colleagues' missed AHT study,[8] physicians continue to miss AHT--and to order unnecessary child-abuse workups. As described in preliminary studies, we estimate that more than 1 in 10 AHT victims admitted for intensive care are still missed or misdiagnosed and that 2 of every 3 patients with accidental head trauma undergo an unnecessary workup for abuse.[25]

A2. To improve the accuracy of current AHT screening and evaluation practices, physicians need an effective screening tool for AHT. An effective AHT screening tool (1) is simple, safe, readily available, highly reliable, inexpensive, and evidence-based; (2) will perform with levels of sensitivity and specificity higher than achieved in current AHT screening practices; (3) has been broadly validated in different geographic regions, in large and small PICUs, and in the same institutions over time; (4) can be applied at or near the time of PICU admission; (5) will facilitate simple, rapid bedside calculation of an evidence-based patient-specific estimate of the probability of abuse; and (6) will provide a clear recommendation to complete an abuse evaluation if/when appropriate to do so. Finally, in the absence of a gold standard, an effective AHT screening tool must perform well across a broad spectrum of criteria physicians use to diagnose abuse.

A3. Unfortunately, an AHT screening tool with all of these qualities did not exist—until now. The PI's team conceptualized, designed, and directed sequential multicenter studies of 500 acutely head-injured children across 18 participating sites to derive and validate an effective AHT screening tool.[22, 23] This new AHT screening tool is simple, safe, readily available, highly reliable, inexpensive, and evidence-based. Its AHT screening performance has been broadly validated in different geographic regions, disparate clinical settings, and against a broad spectrum of definitional criteria for AHT.[23] Findings revealed that, <u>applied accurately and consistently, it would have correctly identified (categorized as high risk) 96% of patients who met our original criteria for AHT, 98% of patients ultimately diagnosed with AHT, and 99% of patients whose completed skeletal surveys and/or retinal exams revealed corroborating findings of abuse</u>.[22, 23] Based on these strong preliminary studies, the AHT screening tool presented here is fully ready for a clinical trial in active PICU settings.

A4. The new AHT screening tool comes in the form of a CDR. A clinical prediction rule is an evidence-based tool that measures and then combines the specific predictive contributions of multiple clinical findings or test results to estimate or predict the probability of a diagnosis, prognosis, or response to therapy in an

individual patient.[34] A clinical prediction rule rises to the level of a clinical decision rule (CDR) if and when physicians use the tool to guide a specific clinical decision. Our screening tool was developed to serve first and foremost as an effective decision rule. Our rule was derived and validated in strict accordance with established guidelines[35-39] and was developed to function effectively as an AHT screening tool in PICU settings.[22, 23]. It includes four highly reliable predictors (see Table A1) and makes a recommendation to complete an abuse evaluation on all patients that it categorizes as high risk (see Table 1). Because the CDR's predictor variables are all readily available, physicians can apply this AHT screening tool when they need it—at or near the time of PICU

Table A1. Validated AHT Clinical Decision Rule
Every acutely head-injured infant or young child hospitalized for intensive care* presenting with **one or more** of four signs and symptoms should be considered high risk and thoroughly evaluated for abuse.
1. Any clinically significant respiratory compromise at the scene of injury, during transport, in the emergency department, or prior to admission
2. Any bruising involving the child's ear(s), neck, or torso
3. Any subdural hemorrhage(s) or fluid collection(s) that are bilateral OR involve the interhemispheric space
4. Any skull fracture(s) other than an isolated unilateral non-diastatic linear parietal skull fracture
*Excluding victims of motor vehicle accidents or collisions.

admission—to inform or guide their early decisions to either launch or forgo child-abuse evaluations in their young, acutely head-injured patients.[22, 23] Physicians will be able to apply this validated AHT screening tool as a clinical decision rule.[22-25] If applied as recommended, it will categorize all patients who present for PICU admission with any one or more of its four variables as high risk and make a recommendation to evaluate all high risk patients for abuse.[23, 25]

A5. Applied accurately and consistently, the CDR could have widespread beneficial impacts. As summarized in Preliminary Studies (section C), accurate and consistent application of this new, broadly validated CDR could: (1) reduce disparities in current AHT screening and evaluation practices; (2) minimize the negative impacts of clinicians' inexperience, uncertainty, and inherent biases; (3) significantly increase AHT detection (87% to 96%); (4) increase the overall diagnostic yield of patients' completed abuse evaluations (i.e., the percentage of patients whose completed skeletal surveys and/or retinal exams reveal corroborating findings of abuse; 49% to 56%); (5) lower evidence-based estimates of missed AHT among patients not evaluated for abuse (0.19 to 0.07); (6) decrease unnecessary abuse evaluations of patients with accidental head trauma (67% to 60%); (7) lower the cost per correctly identified child with AHT (by 15.1%); and (8) dramatically lower overall AHT-associated acute health care costs in PICU settings (by 74.2%).[25] Most importantly, the CDR will save lives—the lives of children who will otherwise suffer fatal inflicted injuries because their AHT is missed or unrecognized.[8, 25]

B. INNOVATION

B1. The CDR addresses a high-stakes clinical decision that no other AHT screening tool addresses. In the realm of pediatric AHT, only two other CDRs are currently under development. Maguire and colleagues have derived and validated an assistive prediction rule that measures the association between AHT and various combinations of six clinical findings.[40, 41] Because two of these six findings can only be confirmed (or excluded) after an abuse evaluation has been launched, their prediction rule cannot be applied as an AHT screening tool at or near the time of PICU admission, when physicians need evidence to inform or guide their early decisions to launch or forgo an abuse evaluation. Berger and colleagues are working to validate a more directive decision rule designed to identify infants at increased risk of brain injury or AHT who might benefit from neuroimaging in the emergency department.[42] These two rules and our CDR target three different decision points along the continuum of AHT clinical care.

B2. The CDR casts a broad screening net to miss as few cases of AHT as possible. The CDR demonstrated sensitivity of 0.96 and specificity of 0.43 in our completed validation study.[23] That is, the CDR would have categorized 96% of AHT patients as high risk and 43% of non-AHT patients as low risk. Positive and negative predictive values were 0.55 and 0.93, respectively. In other words, the proportion of high-risk patients with AHT was 55%, and the proportion of low-risk patients with non-AHT was 93%. The CDR was designed to be highly sensitive to high-risk patients in need of evaluation. This includes opting for lower specificity in favor of higher sensitivity to be liberal in our inclusion of patients at high risk.

B3. The CDR's patient-specific estimates of abuse probability are good predictors of the results of patients' completed abuse evaluations. There are 16 combinations of the CDR's four prediction variables, leading to 16 subpopulations. Across these subpopulations, the CDR abuse probability correlated positively and very strongly ($R = 0.71$) with the overall diagnostic yields,[25] that is, those corroborated by completed skeletal surveys and/or retinal exams.

B4. Our proposed CDR clinical trial is itself highly innovative. Our proposed SCRT will be the very first formal clinical trial of a validated child-abuse screening tool. Our matched pair cluster randomized trial design will facilitate the: (1) assessment of the CDR's impact on relevant clinical outcomes; (2) identification of site-

specific, provider-specific, and patient-specific factors that impact the CDR's acceptability, utilization, and accuracy; and (3) assessment of the effectiveness of the planned multifaceted CDR implementation strategies. Finally CDR implementation will result in: (1) higher diagnostic yields (i.e., abuse corroborated by clinical tests), (2) lower evidence-based estimates of "missed AHT" among patients not evaluated for abuse, and (3) lower percentages of patients with accidental head trauma evaluated for abuse.

C. PRELIMINARY STUDIES

Our proposed CDR clinical trial represents the vital third step in the study PI's long-term research plan to develop an effective PICU-based AHT screening tool. The PI's commitment to the development and execution of this long-term research plan began when he and Dr. Jenny met the infant with missed AHT who would became the inspiration for their sentinel study of missed AHT.[8] This program of work required the formation of a dedicated research network, the application of established methodologies for CDR development, and the successful execution of sequential multicenter studies to derive and validate an effective AHT screening tool.[22,23] Our preliminary studies include the results of our prospective multicenter CDR derivation and validation studies and secondary analyses designed to: (1) further validate the CDR's screening performance; (2) assess site-by-site disparities in current AHT screening and evaluation practices; (3) demonstrate the CDR's application as an assistive prediction tool; (4) estimate the magnitude of the CDR's potential clinical impacts and cost benefit; (5) identify factors that could impact the CDR's acceptability, utilization, and accuracy when applied in PICU settings; and (6) estimate baseline CDR acceptability among potential physician users.

C1. CDR derivation and validation studies. These two prospective cross-sectional observational studies[22,23] were conducted in accordance with established standards for CDR development[35-39] at 18 sites in the United States and Canada. Ten sites participated in both studies. In both studies, clinicians were kept blinded to the CDR's final form and content. The two studies used the same inclusion and exclusion criteria, data forms, methods, and *a priori* definitional criteria for AHT. To minimize circular reasoning, these criteria contained no references to any of the clinical variables considered for inclusion in the CDR. Eligible patients were children less than three years old who were hospitalized for intensive care of symptomatic, acute, closed, traumatic, cranial, or intracranial injuries confirmed on initial head CT or MRI. Patients were excluded if their head injuries resulted from collisions involving motor vehicles or if initial neuroimaging revealed clear evidence of pre-existing brain malformation, disease, infection, or hypoxia-ischemia.

The CDR derivation study.[22] As shown in Table C1, we conducted the derivation study between February 2010 and August 2011 at 14 PICU sites in order to identify a cluster of "predictor variables" that would perform effectively as an AHT screening tool. Our derivation study (1) captured historical, clinical, and radiological data for 209 eligible patients; (2) identified 13 highly discriminating ($p \leq 0.001$) and highly reliable (κ = 0.68-0.95) clinical variables readily available at or near the time of PICU admission; (3) calculated each variable's isolated predictive qualities (i.e., sensitivity, specificity, predictive values, likelihood ratios, post-test probabilities); (4) applied a recursive partitioning algorithm designed to penalize missed cases of AHT; and (5) identified four clusters of these 13 variables that, alone or in combination, detected AHT with a sensitivity $\geq$ 0.92. These four clusters of variables became our four candidate CDRs.

The CDR validation study.[23] To verify the candidate CDRs' AHT screening performance in a new, equivalent patient population, we conducted a CDR validation study between March 2012 and July 2013 at 14 participating sites. Our validation study (1) captured equivalent historical, clinical, and radiological data for 291

Table C1. CDR Performance in the Derivation and Validation Studies					
Derivation Study (N=209)			**Validation Study (N=291)**		
Applying 4-variable CDR	Applying a priori AHT criteria		**Applying 4-variable CDR**	Applying a priori AHT criteria	
	AHT	Non-AHT		AHT	Non-AHT
Higher Risk	91	73	**Higher Risk**	119	96
Lower Risk	4	41	**Lower Risk**	5	71
	Value	(95% CI)		Value	(95% CI)
SENS	0.96	(0.89,0.99)	**SENS**	0.96	(0.90,0.99)
SPEC	0.36	(0.27,0.46)	**SPEC**	0.43	(0.35,0.50)
PREV	0.45	(0.39,0.52)	**PREV**	0.43	(0.37,0.49)
PPV	0.55	(0.48,0.63)	**PPV**	0.55	(0.48,0.62)
NPV	0.91	(0.78,0.97)	**NPV**	0.93	(0.85,0.98)

Abbreviations: AHT=abusive head trauma, CDR=clinical decision rule, CI=confidence interval, PPV=positive predictive value, PREV=prevalence, NPV=negative predictive value, SENS=sensitivity, SPEC=specificity.

additional eligible patients, (2) confirmed AHT screening performance of all four candidate CDRs in a new equivalent patient population, and (3) determined that a four-variable CDR (see Table A1) demonstrated the optimal combination of simplicity and AHT screening performance. All measures of this four-variable CDR's screening performance matched or exceeded equivalent results from the completed derivation study. Most importantly, the CDR demonstrated sensitivity of 0.96 in both studies (see Table C1).[22, 23] As Table C1 shows, 379 (76%) of these 500 eligible patients would have been categorized as high risk.

C2. Secondary analyses. We conducted multiple secondary analyses of the captured data regarding the 500 acutely head-injured patients in our derivation and validation study datasets.

CDR's abuse probability can be used to aid in investigative and legal responses to suspected AHT. For

example, we showed that the mean probability of children reported to child protective services was 0.64 whereas the mean probability of children not reported was 0.20.

The CDR has been broadly validated. Analysis of the CDR's AHT screening performance in novel patient cohorts revealed that the CDR performed with high sensitivity (0.95 to 0.98) in different geographical regions, in large and small PICUs (0.94 to 0.99), and in the same PICUs over time (0.97 to 0.99).[23]

The CDR is highly sensitive no matter how we define AHT. Gold standard definitional criteria for AHT do not exist. As defined by our original *a priori* criteria, the prevalence of AHT in our combined patient population was 0.44. The CDR would have correctly identified (i.e., categorized as high risk) 96% of patients who met these criteria. Applying several modified (more inclusive and less inclusive) AHT criteria to iteratively re-sort the patient population, our resulting calculations of AHT prevalence ranged from 0.28 to 0.52. Subsequent analyses revealed that the CDR would have correctly identified more than 96% of AHT patients no matter how we defined AHT. The CDR would also have correctly identified 98% of patients ultimately diagnosed with probable or definitive AHT by their treating and/or consulting physicians, and 99% of patients whose completed skeletal surveys and/or retinal exams revealed corroborating findings of abuse.[23]

Current AHT screening and evaluation practices demonstrate site-by-site disparities. The percentage of eligible patients (*N* = 500) who underwent both skeletal survey and retinal examination varied from 33% to 100% across the 18 participating sites. The percentage of eligible patients who underwent at least one of these abuse evaluations varied from 43% to 100%, and the percentage who underwent neither abuse evaluation varied from 0% to 57%.

Applied as an assistive prediction tool (i.e., the AHT Probability Calculator), the CDR facilitates the simple calculation of an evidence-based, patient-specific estimate of abuse probability. To calculate patient-specific evidence-based estimates of abuse probability, we (1) divided the patient population (*N* = 500) into its 16 subpopulations defined by patients' unique combinations of the CDR's four predictor variables, (2) calculated the percentage of patients within each subpopulation who met our *a priori* definitional criteria for AHT, and (3) assigned that same value—expressed as an estimate of

Table C2. Extrapolated Measures of AHT screening accuracy in…	Current PICU Screening for AHT	AHT Screening Guided by the CPR	% change
…AHT detection	191 (87%) of 220	212 (96%) of 220	↑ 9%
…Overall diagnostic yields	191 (49%) of 391	212 (56%) of 379	↑ 7%
…Abuse probability among patients not evaluated thoroughly for abuse	0.19	0.07	↓ 12%

abuse probability—to every patient within that same subpopulation. The resulting patient-specific estimates of abuse probability varied widely (from 0.06 to 1.00) across the 16 patient subpopulations. For the 121 low-risk patients (who presented with none of the CDR's four predictor variables), their estimated probability of abuse was 0.07 (95% CI: 0.04-0.14).[25] As discussed above, the CDR's estimates of abuse probability correlated positively with the overall diagnostic yields of patients' completed skeletal surveys and/or retinal exams.[25]

The CDR will improve AHT screening accuracy. In our completed CDR derivation and validation studies, PICU Providers and child abuse consultants completed abuse evaluations on only 277 (**73%**) of 379 patients the CDR would have categorized as high risk, and completed abuse evaluation on 70 (**58%**) of 121 patients the CDR would have categorized as low risk. It is on these preliminary findings that we base the power calculations of the proposed SCRT. These observations led us to complete analyses designed to estimate the CDR's potential impact on AHT screening accuracy.[25] To do so, we applied mean patient-specific estimates of abuse probability to predict additional positive abuse evaluations (i.e., cases of missed AHT) among subgroups of patients lacking skeletal surveys and/or retinal exams and then used these predictions to extrapolate and compare AHT detection in "current PICU screening for AHT" versus "AHT screening guided accurately and consistently by the CPR." Our results (see Table C2) suggest that the CDR could (1) increase AHT detection (from 87% to 96%), (2) increase the overall diagnostic yield of patients' completed abuse evaluations (from 49% to 56%), (3) reduce (mean) evidence-based estimates of abuse probability (i.e., estimates of missed AHT) among patients who are not evaluated for abuse (from 0.19 to 0.07), and (4) reduce unnecessary abuse evaluations of patients with non-AHT (from 67% to 60%).[25]

The CDR could also decrease AHT-associated acute health care costs in PICU settings. Applying average skeletal survey, retinal exam, and visit costs obtained from the MarketScan Commercial Claims and Encounters database, we estimate that accurate and consistent application of the CDR in PICU settings could reduce the average cost per correctly identified child with AHT from $458 to $389—a reduction of 15.1%. Applying published data on re-injury rates in children with missed AHT[8] and the costs of emergency department/clinic visits and hospital re-admissions resulting from re-injury[9], we estimate a 72.4% reduction in total AHT-associated acute health care costs stemming from the anticipated reduction in missed AHT cases.

Non-clinical factors appear to impact physicians' decisions to launch or forgo child-abuse evaluations in

their young acutely head-injured patients. Exploratory analyses identified four non-clinical variables associated with significant differences ($p < .05$) in the percentages of high- or low-risk patients evaluated for child abuse. These include primary caregiver denial of any head trauma; patient of minority race/ethnicity; patient cruising or walking prior to PICU admission; and patient age less than 6 months at time of PICU admission.

A preliminary assessment of CDR acceptability yielded favorable results. The Ottawa Acceptability of Decision Rules Instrument (OADRI) is a validated 12-item instrument that evaluates decision rule acceptability among clinicians.[43] The OADRI's final score is calculated as the mean of all 12 items, resulting in a score that ranges from 0 to 6 with 6 being greatest acceptability. To estimate the CDR's baseline acceptability among potential users, we distributed a brief, voluntary, anonymous survey to pediatric intensivists and child-abuse pediatricians who participate in the dedicated listservs of the PALISI research network and the Ray E. Helfer Society. Survey participants were provided a one-page document that summarized the CDR's derivation, validation, and potential clinical impact.[22, 23, 25] Participants were asked to answer the OADRI's 12 questions "as if they were considering using the rule." A total of 30 pediatric intensivists, 51 child-abuse pediatricians, and 2 other clinicians completed the survey. The mean OADRI score for the total sample (N= 83) was 4.86 (SD =1.32). These participants endorsed items indicating that the CDR was "easy to use", "easy to understand", "useful", "clear", "supported by colleagues", and "will benefit patients". Though not generalizable, these results suggest that clinicians will consider the CDR's adoption as an AHT screening tool in PICU settings.

Summary. The PI's (Hymel) preliminary studies clearly demonstrate that CDR is simple, safe, inexpensive, reliable, evidence-based, and highly sensitive. It was derived and validated in strict accordance with established standards for CDR development. Secondary analyses demonstrate its clear potential to improve patient safety (i.e., to reduce the likelihood that patients with AHT go undetected) and clinical efficiency (i.e., to reduce unnecessary abuse evaluations and their associated costs). These preliminary studies also demonstrate that roughly 76% of head emergency room trauma cases are considered high risk, but only 73% of these high-risk cases receive an evaluation for abuse. To enhance clinical practice, the CDR needs (and is now ready for) a formal clinical trial. To conduct this trial we have assembled an extraordinary team. Dr. Chinchilli (Co-I)—an experienced biostatistician with a broad background in clinical trials, longitudinal data analysis, multivariate analysis, crossover trials, and measures of agreement. Dr. Wang (Co-I)—a biostatistician with expertise in analyses of large-scale longitudinal studies—has collaborated with the Study PI (Hymel) as well as other clinical investigators to develop and implement advanced statistical methodologies. Finally, Dr. Dias (Co-I)—a pediatric neurosurgeon recognized internationally for his AHT clinical expertise and his research on AHT prevention will apply his vast clinical experience in monitoring the long-term care and outcomes of a cohort of head-injured patients whose clinical data will contribute to the TCCMS Admin Core's data repository.
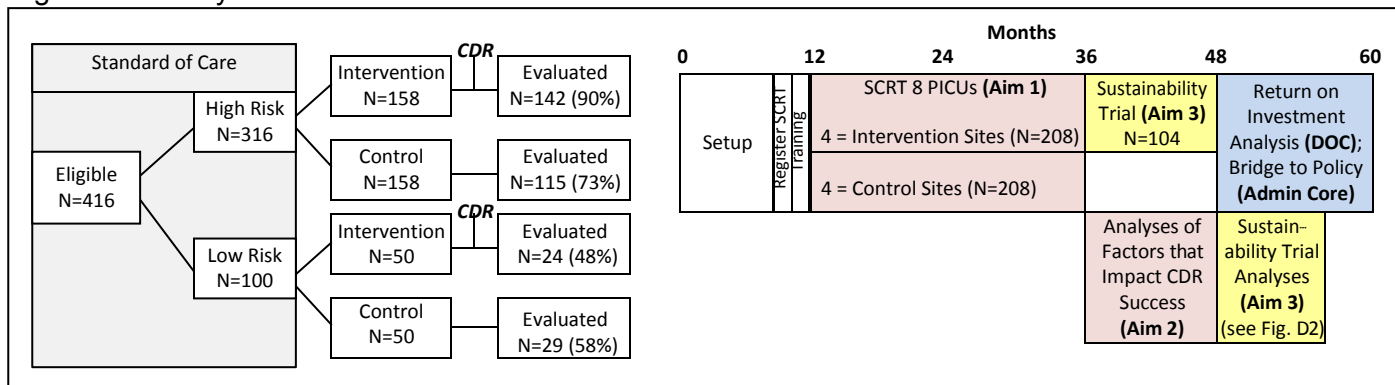
D. APPROACH

D1. Overview. To assess the CDR's impact on AHT screening accuracy, we propose a stratified cluster randomized trial (SCRT) at eight university-based US PICUs to test "AHT screening guided by a CDR" compared to "AHT screening as usual." Participating sites will include Connecticut Children's Medical Center, Hartford; Virginia Commonwealth University Medical Center, Richmond, VA; University of Nebraska Medical Center, Omaha, NE; The Children's Mercy Hospital, Kansas, MO; Baylor College of Medicine, Houston, TX; University of Texas Health Science Center at San Antonio, San Antonio, TX; University of Utah, Salt Lake City, UT; and University of Kansas Medical Center, Wichita, KS. The eight PICU sites, located in diverse regions of the U.S. and representative of large and small PICU settings, will be matched into four pairs based on each site's projected volume of high-risk patients. One PICU from each pair (4 total) will then be randomized to the intervention arm of the SCRT. The remaining 4 PICUs will serve as controls. Randomization sequences will be kept by the study biostatistician (Dr. Chinchilli) and will be concealed from Dr. Wong, who will be blinded to group membership and responsible for formal data analyses of Aims 1 and 2.[44] In this SCRT sites will be stratified by patient volume (number of patients seen for AHT via PICU census): there will be four high-volume sites, two of which will be assigned to the intervention and two to control, and four low-volume sites, also equally distributed into the intervention and control groups. Further, in this trial, patients are clustered within Providers which are clustered within sites and thus our power analysis and analysis plan are based on this clustering. Dr. Chinchilli will use a pseudorandom number generator in the statistical software R to obtain the randomization schedule within each stratification. Based on patient volume reports at each of the PICUs, we anticipate 416 eligible patients over a 24-month period (see Figure D1). If eligible patient enrollments are low, we will extend this period by 6 months. Our preliminary studies (Table C2) indicate that ~76% of eligible cases will be deemed high risk based on "standard of care" (N=316). Based on our preliminary studies, we anticipate that only 73% of these high risk cases will receive an evaluation for abuse (N=115). Our conservative estimate (see section D3) is that we will increase this evaluation rate to 90% by implementing the CDR (N=142).

Study personnel and coordination. Project 2 involves the coordination of personnel at the Penn State Hershey College of Medicine (COM) and at the 8 participating PICU sites. At the COM, the PI (Dr. Hymel) will oversee all activities. Dr. Hymel will supervise the Overall Project Coordinator (OPC) at the COM and will work closely with the biostatisticians. The OPC will be responsible for communicating with all PICU site research personnel, compiling all data in REDCap, working with Dr. Wong on data management, generating site-specific feedback and the monthly booster emails (see section D3), and working with site research personnel at the end of trial to ensure all data are completed.

Each PICU site includes a Site PI, Site Research Coordinator (SRC), and Providers who are defined as PICU physicians and child abuse consultants. At each participating PICU, the Site PI will be responsible for reviewing and communicating site-specific feedback and conducting information-sharing sessions (see section D3). The SRCs will be responsible for ensuring data is compiled. The data includes the information from the data forms (Appendix B2) and information gleaned from the EMR (see Table D1). Data form and EMR information is transferred to REDCap by the SRCs. Providers are the participants of the clinical trial. It is these Providers who will apply the CDR, consider its recommendations, and make decisions to either launch or forgo child-abuse evaluations. It is their reasoning and decision-making behavior that will be captured on the data forms by the SRC and transferred to REDCap for analysis.

Clinical trial. As outlined in Figure D1, the first year will consist of project set-up and training of the implementation strategies. Prior to full implementation (during Months 10–12), Providers at the intervention sites will undergo online training to learn about the screening tool and its application, and Providers at both the intervention and control sites will gain familiarity with the study design, methods, data forms, and procedures. The SCRT will commence by the end of Month 12. Over the next two years (Months 12–48), active strategies (D2.2) will be deployed at the intervention sites to promote PICU Providers' adoption of the CDR. These Providers will complete four measures of CDR acceptability (OADRI, see Appendix B6) at six-month intervals during the trial. In addition, Site Research Coordinators will work with Providers to ensure that all patient-specific (both the intervention and control sites) and provider-specific (only the intervention sites) data are continuously captured throughout the trial.

Figure D1. Study flow and timeline

Using these data, we will compare the number of high-risk and low-risk patients who are actually evaluated for child abuse (i.e., patients who undergo both skeletal survey and retinal exam) at the intervention and control sites to assess the CDR's impact on AHT screening accuracy (Aim 1). Beginning in Month 36, we will conduct analyses using data collected during Months 12-36 (see Appendix B2) to determine whether site-specific, provider-specific, and/or patient-specific factors impact AHT screening accuracy (Aim 2). Finally, to address our Exploratory Aim regarding the sustained impact of our implementation efforts, we will measure CDR utilization at the intervention sites over the 12-month period following the discontinuation of active CDR implementation strategies (Exploratory Aim 3).

D2. Participants. To test our hypotheses with sufficient power, we need to capture the historical, clinical, and radiological data for 416 eligible acutely head-injured infants and young children. Replicating the criteria used in our completed CDR derivation and validation studies,[22, 23] eligible patients will be children under three years of age admitted to the PICU for the management of symptomatic, acute, closed, traumatic, cranial, or intracranial injuries confirmed by computed tomography (CT) or magnetic resonance imaging (MRI). Patients will be excluded if their acute head injuries resulted from a collision involving a motor vehicle or if initial neuroimaging reveals clear evidence of pre-existing brain malformation, disease, infection, or hypoxia-ischemia. Eligible patients will be classified as high or low risk. A high-risk patient is defined as presenting to the PICU with closed head trauma and at least one of the four following conditions: (1) acute respiratory compromise; (2) bruising of the torso, ear(s), or neck; (3) bilateral or interhemispheric subdural hemorrhage(s)

or fluid collection(s); and (4) any skull fracture(s) other than an isolated unilateral non-diastatic linear parietal skull fracture. A low-risk patient will have none of these presenting conditions. In our preliminary studies, Providers completed abuse evaluations on only 73% of their high-risk patients (see Section D1). Based on census estimates from participating PICUs, we expect 304 high-risk patients and 112 low-risk patients (see Figure D1). Based on the census from participating Pediatric Intensive Care Units, we anticipate the patients being 74% White, 17% African American, 9% other with 26% of Hispanic ethnicity. We anticipate that the gender distribution of the patients will be 48% females.

Although we need complete data for 416 patients to test our hypotheses, the actual subjects of this trial will be the Providers (as defined in section D1). We anticipate that participating sites will obtain waivers of parental consent related to capturing patient-related data (see Human Subjects). Only Providers will be required to provide informed consent for study participation. Based on current and projected staffing compositions of, we anticipate an average of 5 Providers at each low patient-volume PICU and an average of 15 for each high patient-volume PICU. Hence we expect to consent 80 Providers across the 8 participating sites.

D3. Intervention. Getting physicians to adopt novel best practices is difficult[45] and usually requires a flexible, multifaceted strategy that identifies and targets local barriers to practice change.[45-58] Adopting an analogous plan to facilitate practice change, we have prepared a cadre of active and adaptive CDR implementation strategies designed to promote CDR adoption as an AHT screening tool at the intervention sites. Directed specifically to Providers at the intervention sites, these strategies will include: (1) initial online training (Appendix B1) regarding the CDR's derivation, validation, application as an AHT screening tool, and potential beneficial impacts which will be augmented with onsite visits by PI Hymel; (2) monthly booster emails designed to reinforce and sustain the effectiveness of the initial online training (Appendix B5); (3) Provider reports of their application of the CDR and active consideration of its recommendations for each eligible new patient admission to the PICU—e.g., "Which of the following explain your current reluctance to evaluate your 'higher risk' patient for abuse?" (see Appendix B2); and (4) ongoing Provider access to a fully developed online or smartphone "AHT Probability Calculator" app (Appendix B7) that will facilitate the simple, rapid bedside estimate of abuse probability and provide real-time data capture regarding use and probability of each case considered. More specifically, the calculator predicts the likelihood that a child's head injury was caused by abuse based on the 16 combinations of the CDR's four prediction variables (see Section B3). Each of the 16 combinations has a specific, associated probability of abuse, which the calculator displays to help guide Providers' decisions. Finally, (5) Providers will carry a CDR badge card listing the four CDR prediction variables (see Table A1).

Online training. The initial online computer-based training will include background information such as statistics on AHT, the four variables included in the CDR, how/when to apply the CDR, and beneficial impacts of CDR application. A series of scenarios of high-risk and low-risk cases is presented; Providers use the AHT probability calculator to derive an estimate of abuse probability and then chose whether or not to further evaluate the child for abuse. Of the first 6 scenarios, if 5 out of 6 (83%) cases are correctly identified as needing/not needing evaluation, the training evaluation is completed. If the 83% criterion is not achieved, additional scenarios are presented until a score of 83% is achieved (Appendix B1). The online training will be available on a secure website created for the study. Providers will also receive a CDR badge for quick access to the CDR as another aid for screening of potential abuse. The PI (Hymel) will conduct on-site visits with Providers to train research coordinators, augment the online training, and facilitate conversation on barriers to implementation. This training has been piloted with four experienced child abuse pediatricians who have reviewed the content, layout, and structure.

Site-specific feedback. To identify and address local barriers to practice change, site PIs will receive iterative site-specific feedback via email regarding their Providers' AHT screening performance over the prior six months (see example, Appendix B4). This feedback will be provided for four (six-month) cycles during the trial's 24 months of active data capture and will include the following information: (1) a list (in descending order of frequency) of local Providers' self-reported barriers to CDR utilization, accuracy, and/or acceptability (e.g., "Which of the following explain your current reluctance to evaluate your "higher risk" patient for abuse?"—see Appendix B2); (2) Providers' most mean score on the OADRI (see Section D1 above); and (3) graphs that compare outcome measures of CDR utilization across the four intervention sites over the previous six months. Upon receipt of site-specific feedback, site PIs will schedule and facilitate a local information-sharing session so their Providers can review the feedback and discuss strategies for overcoming local barriers to CDR utilization, accuracy, and acceptability.

Tailored booster emails to Providers. Monthly emails will be sent to Providers that will include feedback indicating their CDR utilization compared to the other intervention site Providers; a recommendation, statistic,

or reminder from the online training; a recent or relevant research article; and an invitation to ask questions or provide comments.

CDR Badge Card. Physicians will receive a CDR badge card listing the four CDR prediction variables (Table A1). The badge will allow another avenue for a quick screening of AHT and can be used in conjunction with the AHT Probability Calculator when a physician is unsure whether a patient is at high versus low risk and need for evaluation. This CDR card will affix to their ID badge for quick and easy access.

AHT Probability Calculator. Providers will download a smartphone or mobile-device "AHT Probability Calculator" (Appendix B7) that will facilitate simple, rapid bedside estimates of abuse probability. Based on our preliminary studies, the calculator predicts the likelihood that a child's head injury was caused by abuse based on the 16 possible combinations of the CDR's four prediction variables (Section B3). Each of the 16 combinations has a specific, associated probability of abuse, which the calculator displays to help guide Providers' decisions. When calculators are used, data will be captured via the apps statistics for later analyses.

Adherence. Strict adherence to the CDR's recommendations will be encouraged—but not required—for several reasons: First, the CDR is imperfect—it will categorize some patients with non-abusive head trauma as high risk and will categorize a few patients with AHT as low risk. Second, Providers' clinical judgment will sometimes conflict with the CDR's categorization. Third, many Providers will not agreed to completely bypass their clinical judgement in favor of strict adherence to the CDR's categorizations (high versus low risk) or recommendations. Finally, individual and/or institutional barriers will likely limit the CDR's accurate and consistent application. *To be conservative, we predict the CDR implementation will motivate intervention site Providers to complete abuse evaluations on 90% of their high-risk patients during this SCRT, and we have thus powered Aim 1 to detect a 17% increase in high-risk patients being screened (i.e., 73% to 90%).*

D4. Data capture to evaluate CDR utilization. Under current "standard of care" practice, patients presenting with head trauma are identified through daily screening of PICU admission logs for initial risk level (see Figure D1). The CDR will be applied where high-risk patients should receive a child abuse evaluation. A completed skeletal survey and a completed retinal

Table D1. Variables used in Analyses: Informant and Source

	Variable	Definition	Source
Aim 1	Categorization of ↑ and ↓ risk patients	"Standard of Care" Admitting Criteria: high risk=presenting with closed head trauma + any of four CDR variables. Low risk=closed head trauma with none of the four CDR variables.	EMR (Table A1)
	% of ↑ risk and ↓ risk patients evaluated for abuse	Abuse evaluation is defined as provider orders skeletal survey and retinal exam	EMR
	Diagnostic yield	Corroborating findings from skeletal survey and retinal exams	EMR
	% missed AHT among patients lacking abuse evaluation	The number of patients lacking abuse evaluation with ↑probability as determine by the AHT probability calculator	EMR & AHT Calculator (Section D3)
Aim 2	Site-specific: High vs. low patient volume	Hospital census of PICU admittance	EMR
	Provider-specific: Child abuse expertise	Provider Credentials (Q3.2.1 & Q3.2.2)	Data form Appendix B2
	Provider-specific: Exposure to implementation	Points received for participating in implementation strategies	Various* (see Table D2)
	Provider-specific: CDR acceptability	Provider self-report on OADRI every 6 months	OADRI
	Patient-specific: Age	Age of patient	EMR
	Patient-specific: Race	Race of patient	EMR
	Patient-specific: Caregiver denial of AHT	Queried by provider during initial intake (Q5.2.3)	Data form Appendix B2
	Patient-specific: Mobility (walking/cruise)	Observation by provider during initial intake (Q5.2.6)	Data form Appendix B2
Aim 3	% of patients evaluated for abuse	Abuse eval=provider orders skeletal survey & retinal exam	EMR
	CDR Acceptability	Provider self-report on OADRI every 6 months	OADRI
DOC	Length of Stay	For return on investment calculation for DOC Aim 1	EMR (Section F)

*Data form, AHT Calculator, Email Receipts, Training Results, OADRI, Site PI Reports

exam constitute a child abuse evaluation. At both intervention and control sites, Research Coordinators will record whether or not patients receive a subsequent child abuse evaluation (see Figure D1 and Appendices 4-6); all data will be entered via REDCap on a secure server. Table D1 summarizes the variables used in all aims, how they are defined, and the sources of each. As outlined above, Site Research Coordinators will be responsible for extracting EMR data and completing all data forms in REDCap.

D5. Measures for Aim 1. The two predicted results relevant to Aim 1 include: (1) a higher percentage of high risk patients evaluated for abuse via a skeletal survey and/or retinal exam; and (2) a lower percentage of low risk patients evaluated for abuse. To assess these outcomes, we will calculate the number of patients categorized as either high or low risk based on admitting criteria and will capture data throughout the active trial (Months 12–48) regarding completed skeletal surveys and completed retinal exams. Although we are not specifically powered to detect differences in the following clinical outcome measures, we will nonetheless calculate these as indicators of the clinical impact of the CDR: (1) the overall diagnostic yield of patients' completed abuse evaluations, and (2) the mean estimate of missed AHT among patients lacking evaluations for abuse. We will obtain information for each of these indicators from data collected from Providers during the active trial (Months 12–48) using the forms found in Appendix B2.

D6. Measures for Aim 2. For Aim 2, we will determine what—if any—site-, provider-, and patient-specific factors impact the percentage of high-risk patients who are evaluated for AHT and the percentage of low-risk patients who are evaluated for AHT. Site-specific variables. To determine high and low patient volumes, we will

calculate the number of eligible patients (described in D2) who pass through each intervention site to determine which sites had the highest number of eligible patients and which had the lowest number of eligible patients (based on information from the data-collection forms found in Appendix B2). We will then combine screening-accuracy data from the two sites with the highest patient volumes and the two sites with the lowest patient volumes and will use those data to make comparisons between the two groups. Provider-specific variables. For each case of AHT that arrives in the PICU, the provider will report (using the forms found in Appendix B2) who consulted on case: a standard Provider (without child-abuse expertise), a Provider with child-abuse expertise, or both. As a measure of training "dosage," we will also capture the Providers' exposure to the CDR implementation strategies (see Table D2). Providers will receive points when they complete their initial online training, read the monthly booster emails, apply the CDR and actively consider its recommendations with each new patient, use the AHT Probability Calculator, refer to their CDR badge card, and attend the information-sharing sessions. In some cases, the data will be calculated automatically (e.g., when a Provider opens an email or completes the online training, both of which can be tracked), whereas the rest of the data will be collected using forms found in Appendix B2 (e.g., "Which of the following explain your current reluctance to evaluate your 'higher risk' patient for abuse?"). Finally, acceptability will be assessed every six months and quantified via scores on the OADRI (see C2.8), Using these data, we will compare screening accuracy among Providers to determine the effect of each variable on our two outcomes.

Patient-specific variables. The age and race/ethnicity of patients will recorded as will information on whether caregivers specifically deny head trauma and whether patients have begun to walk/cruise given findings from our preliminary studies (Appendix B3). We will conduct group comparisons to determine what—if any—patient-specific factors are linked to the two outcomes of interest.
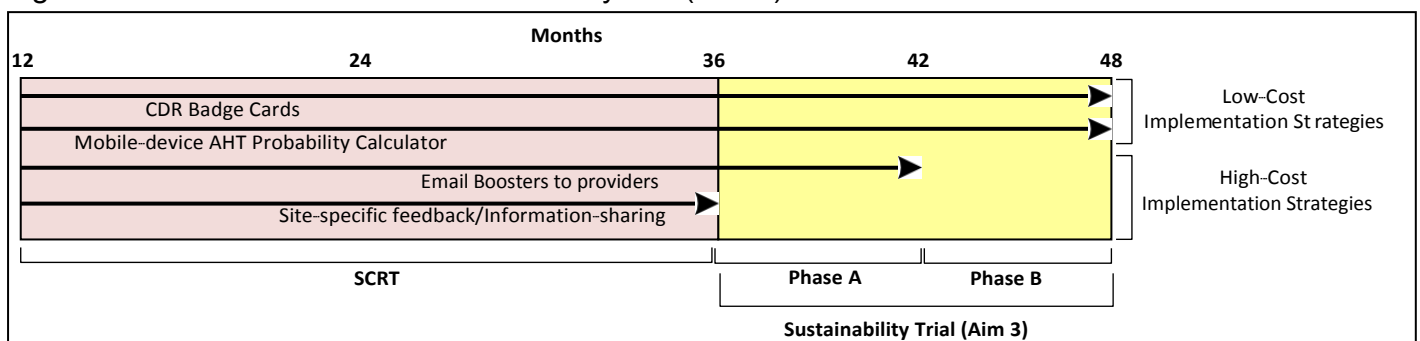
D7. Measures for Exploratory Aim 3. We will conduct a sustainability trial at the implementation sites once the SCRT is completed (at 36 months; see Figure D2). To test the impact of gradually repealing the implementation strategies on sustainability, the local information sessions will be discontinued beginning immediately after the SCRT has ended and the tailored booster emails will no longer be sent to physicians six months after the SCRT has ended. Finally, when these CDR implementation strategies have been discontinued, intervention site Providers will continue to capture patient-specific data and complete a final measure of CDR acceptability. Physicians will complete a final phone call interview focused on barriers to sustainability. The data collection phases of Aim 3 will provide important

Table D2. Exposure to CDR Implementation Strategies

Strategy	Calculated	Points Possible
Initial Online Training Completed	**Completion**: 1 pt for score of 90% or more.	1
Read Monthly Emails (Return Receipt)	**24 emails**: 1 pt per email acknowledged with return receipt.	24
Apply the CDR (from Data Form)	**% used**: 0-60%=1; 61-70%=2; 71-80%=3; 81-90%=4 (Goal); 91-100%=5	5
AHT Probability Calculator (app capture)	**% used**: 25% (1-2 times)=1; 50% (3-4 times)=2; 75% (5-6 times)=3; 100% (>6 times)=4	4
CDR Badge Card (OADRI)	**% used**: Never=1; Rarely, < 20%=2; Occasionally, ~40%=3; Sometimes, ~60%=4; Frequently, ~80%=5; Always, 100%=6	6
Attend Information Sharing Sessions	**Attendance**: 1 pt for each of 5 sessions (Start of trial and every 6 months)	5
	TOTAL PTS	45

Figure D2. Flow and timeline of sustainability trial (Aim 3)

quantitative and qualitative information about the utility of the most resource-consumptive implementation supports for CDR implementation. Phasing out the most consumptive and least utilized supports, without loss in CDR use, makes it more likely that hospitals and physicians are able to sustain implementation. If CDR acceptability and utilization falls, then the importance of these supports will be better understood.

This goal will be to evaluate whether Providers continue to use the CRD when implementation support is reduced. We will systematically reduce high-cost implementation supports one by one and evaluate the impact of these reductions on: (1) CDR use (the percent of high-risk patients evaluated for abuse), and (2) Provider acceptability as measured by the OARDI—both of which will be assessed at 36, 42 and 48 months. As Figure D2 depicts, site-specific feedback will be discontinued at 36 months, and email boosters to Providers will be

discontinued at 48 months. We anticipate that another 104 patients will be eligible during this 12-month period (Figure D1). This nested approach will allow us to compare the impact of site-specific feedback (Phase A) and email boosters (Phase B) to full implementation support (during the SCRT) and to one another. This sustainability trial will employed us to test whether CDR utilization and acceptability can be maintained when only low-cost strategies are used. Further, brief interviews will be conducted with Providers about utilization and acceptability one year after active implementation supports are removed. This strategy will provide key information about the process of sustaining the CDR and cost reduction possibilities in nation-wide dissemination.

D8. Power Analysis. To estimate the statistical power for the target sample size of 304 high-risk patients based on the first primary measure of CDR utilization (see section D5), we simulated 1,000 datasets via the generalized linear mixed-effects model described below by excluding the patient-level covariates under the following set of assumptions: [1] Sites 1, 3, 5, and 7 are randomized to the control group, whereas sites 2, 4, 6, and 8 are randomized to the intervention group; [2] The mean percentages of high-risk patients evaluated for abuse (i.e., skeletal survey and retinal exam) are 73% in the control group and 90% in the intervention group based in part on observations from our completed preliminary derivation and validation studies (section D2). We analyzed each of the 1,000 simulated datasets, each of sample size 304, via SAS PROC GLIMMIX, Version 9.4, with respect to the primary null hypothesis (Aim 1). The estimated statistical power (proportion of simulated datasets that yielded a statistically significant result for Aim 1) was 95.7%. Therefore, our target sample size of 304 high-risk patients has excellent statistical power for addressing Aim 1. For Aim 2 we consider the patient-level factor of minority race and/or ethnicity for power estimation to investigate its significant impact on the percentage of low-risk patients at least partially evaluated for abuse using the intervention group only. Our preliminary studies (section C) indicate that minority patients will have a 1.5 higher log odds of being evaluated for abuse than non-minorities. The estimated power (proportion of simulated data sets that yielded a statistically significant result for Aim 2) was 90% for testing the null hypothesis $H_1: \beta = 0$ against the alternative hypothesis $H_1: \beta = 1.5$.

D9. <u>Data Analyses</u>. *<u>Data Management</u>*. Supervised by Dr. Hymel, the Overall Project Coordinator (OPC) will be responsible for monitoring the data as they are input via REDCap by Site Research Coordinators. The OPC will work closely with Dr. Wong to develop data quality assessment reports, including accrual and attrition reports, which will be shared with the PI and the DSMB. Dr. Wong will be blinded to group membership.

<u>Steps Taken Prior to Conducting Formal Analyses.</u> The primary statistical package used will be SAS, version 9.2. Prior to formal analyses, means, medians, ranges, standard deviations and descriptive measures for nonnormality will be computed for each continuous variable, as well as frequencies for categorical variables, within each group. Any data values identified as outliers will be further examined to determine if they are data entry errors, in which case they will be modified when possible. Outliers will generally be kept in the final analysis, unless substantial evidence is available for their deletion. If the distribution of a continuous outcome variable varies substantially from a normal distribution, a transformation to approximate normality will be performed, and all formal analyses will be carried out with the transformed variable. If groups exhibit heterogeneous variances, an appropriate statistical method will be used to account for this fact. Outcome variables will also be examined over time at the group and the individual level graphically as well as via descriptive statistics stratified on time and group. Furthermore, bivariate relations between outcome variables and potential covariates will be examined to determine if substantial deviations from linearity exist.

<u>Intent-to-Treat Analysis Plan and Plan for Addressing Missing Data.</u> The primary approach for the analyses is intent-to-treat, meaning that Providers will be kept in the group to which they were randomized, regardless of protocol violations or drop-out. Drop out will be defined as a Provider who leaves the site. Every effort will be made to minimize missing data, analyses will be performed to explore plausible missing data mechanisms[59], and predictors of missingness will be investigated. If missing data are substantial, maximum likelihood estimation will be employed where predictor variables associated with missingness will be added to the statistical model. This approach generally provides valid inferences when data are missing at random[60, 61] and is more appropriate than many commonly employed approaches to data imputation[62].

D10. Aim 1 Analyses. We will use a linear-mixed model approach for longitudinal data. We let $Y_{l'\#\$}$ denote the binary response of evaluation for abuse with both skeletal survey and retinal exam for the $l^{!!}$ high risk patient from the $k^{!!}$ Provider within the $j^{!!}$ site of the $i^{!!}$ group, where $i = 0, 1$ (corresponding to the control and intervention group), $j = 1, 2, 3, 4$ (corresponding to site), $k = 1, 2, ..., K_{l'}$ (corresponding to Provider), and $l = 1, 2, ..., L_{l'}$ (corresponding to high risk patient). We construct the following bivariate logit function for

generalized linear mixed-effects models: $log_! \left\{ \frac{!''[!_{!''\#\$} !_!]}{!''[!_{!''\#\$} !_!]} \right\} = \mu_! + u_{!''} + v_{!''\#}$, where $\mu_!$ and $\mu_!$ represent means

on the logit scale for the control and intervention sites from health Providers, u_{ij} is a random effect for the jth site of the ith group, and v_{ijk} is a random effect for the kth Provider within the jth site of the ith group. We assume that the u_{ij}s are independent and identically distributed according to a normal distribution with a mean of zero and a variance of σ_u^2; the v_{ijk}s are independent and identically distributed according to a normal distribution with a mean of zero and a variance of σ_v^2, and the u_{ij} s, the v_{ijk} s are mutually independent. We will apply an iterative maximum likelihood (ML) algorithm to derive the estimates of the parameters via SAS PROC GLIMMIX, Version 9.4. The adjusted relative risk (RR) for comparing the intervention and control groups among health Provider is $RR = \{1 + \exp(-\mu_1)\}/\{1 + \exp(-\mu_0)\}$, and the primary null hypothesis is $H_1: RR = 1$, which is equivalent to $H_1: \mu_0 = \mu_1$, where we will construct a one-df F-test based on the ML estimates. We will apply a variety of sensitivity analyses to determine the robustness of the results based on this primary statistical model. In particular, we will investigate heterogeneity of variance by requiring σ^2 to vary according to the matched pair of sites and ω^2 to vary according to the type of health care Provider.

Of note, the same approach and statistical models described above can be applied to **the second measure in Aim 1** (see D5), where Y_{ijkl} is denoted as the binary response of at least partial evaluation for abuse (i.e., skeletal survey and/or retinal exam) for the lth low-risk patient from the kth Provider within the jth site of the ith group. In addition, with regard to additional indicators of the clinical impact of the CDR (see D5), preliminary investigation by using aggregated data over a 24-month period (Month 12-36) can be conducted via two-sample proportional tests with the significance level of .05.

D11. Aim 2 Analyses. We use the same general model introduced in Aim 1, but specify predictors at various levels, and when data are only applicable to the treatment group, we analyze data within the treatment group only. For example, we expect that sites with higher patient volumes will be associated with higher percentages of CDR application, $\log_1 \left\{ \frac{\pi_{ijkl}}{1 - \pi_{ijkl}} \right\} = \mu + u_{ij} + v_{ijk} + \beta x_{ij}$, where x_{ij} is the yield for the jth site within the ith intervention group. With respect to Provider-level risk factors, we will use the following model to examine the association between proportion of CDR application and the level of child-abuse expertise of the Provider: $\log_1 \left\{ \frac{\pi_{ijk}}{1 - \pi_{ijk}} \right\} = \mu + u_{ij} + v_{ijk} + \beta x_{ijk}$, where x_{ijk} is the level of child-abuse expertise for the kth Provider within the jth site within the ith intervention group. Because data for the Provider-level predictors of high exposure to the CDR implementation strategies and high CDR acceptability will only be available within the treatment group, we will examine each of these factors separately as predictors of the proportion of CDR application using the following model within the treatment group only: $\log_1 \left\{ \frac{\pi_{1jk}}{1 - \pi_{1jk}} \right\} = \mu + u_{1j} + v_{1jk} + \beta x_{1jk}$, where x_{1jk} is either the level of exposure to the CDR implementation strategy or level of CDR acceptability for the kth Provider within the jth site within the intervention group (i.e., group 1). Finally, we will examine the patient-level risk factors of minority status and child age/developmental level (cruise/walk) within the following model: $\log_1 \left\{ \frac{\pi_{ijkl}}{1 - \pi_{ijkl}} \right\} = \mu + u_{ij} + v_{ijk} + \beta x_{ijkl}$, where x_{ijkl} is child age or minority status for the ith patient within the kth Provider within the jth site within the ith intervention group. It is important to note that when examining models that include both intervention groups, we will also explore the possibility of any potential intervention group by factor interactions before moving forward with the models with main effects only. For all Aim 2 models, β is the parameter of interest with the corresponding null hypothesis of $H_1: \beta = 0$.

Also within Aim 2, we will assess CDR acceptability using both averages and longitudinal data based on data at Month 12, 18, 24, 30, and 36 (after the intervention strategies are discontinued for the Providers using the score based on the OADRI, a validated 12-item instrument (see D6). Specifically, we will use a linear-mixed model approach for longitudinal data to explore the possibility of linear change over time using a random-effects model in addition to fitting a linear mixed-effects model that calculates Provider-specific means to evaluate the acceptability of the treatment.

D12. Aim 3 Analyses. Similar to the longitudinal model used in Aim 2, we will use a linear-mixed model approach for longitudinal data to explore the possibility of linear change over time (i.e., from 36 to 42 months for Phase A, and from 42 to 48 months for Phase B; see Figure D2). We will use a random-effects model in addition to fitting a linear mixed-effects model that calculates both CDR acceptability (OARDI scores) and CDR application over time and evaluate whether reducing email boosters and/or site-specific feedback have an impact on sustainability. If there is no linear effect, then we can conclude that reducing these high-cost implantation strategies does not significantly reduce the effects of the intervention over time. If linear effects are significant (and negative), then we can conclude that one or more of the high-cost strategies should be

maintained in order to sustain the effectiveness of the intervention.

E. RIGOR, TRANSPARENCY AND REPRODUCIBILITY

In compliance with NOT-OD-16-011 and TCCMS standards (see Admin Core Aim 1), we have designed Project 2 to enhance scientific rigor by articulating our scientific premise in terms of the need for an evidence-based tool to improve care for AHT children. We have outlined the strengths and weaknesses of published research (including our own) to clearly represent why the proposed SCRT is a sound next step in the CDR process. We also have provided details of the design of the SCRT, its Aims, and the intervention and data collection plans. Our methods include clear exclusionary criteria, a randomized trial, objective tools, and multiple data sources including Providers as well as electronic medical records. One of our statisticians will be responsible for the randomization; the other will be blinded to group membership and will conduct subsequent analyses. Our analysis plan was developed in consultation with clinical trials experts and is state-of-the-art, robust, and unbiased. We have *a priori* hypotheses and have been clear when analyses are exploratory. We will register this SCRT with *clinicaltrials.gov* prior to enrollment and have a committed DSMB and solid plan for monitoring the trial. To enhance transparency and reproducibility, de-identified clinical data from Project 2 will be included in the TCCMS's data repository (within the project's Administrative Core) for future hypothesis-generating research that will further advance the field. Permission will also be sought to re-contact the 520 families involved in both the SCRT and the Sustainability Trial for future TCCMS research studies that will grow new science examining: (1) the exacerbated impact of child abuse on head trauma, and (2) how the fundamentally different types of brain injuries manifest in disparate developmental trajectories. Mark Dias (Project 2 Co-I) will begin crafting such an R01 in Year 4.

F. IMPACT

The consequences of AHT can be grave, including the strong possibility of long-lasting, often permanent neuropsychological, behavioral, emotional, and scholastic difficulties as well as the potential for continued abuse, and possibly death, if AHT is missed or unrecognized. However, as we have described, the decision to evaluate a child for AHT is complicated by myriad Provider-, site-, and patient-specific factors. The goal of this CDR, which is the first child-abuse screening tool to undergo a formal clinical trial, is to help Providers effectively overcome barriers and to quickly, reliably, and easily determine when a head-injured child should be evaluated for abuse, thereby eliminating any hesitations that may be inherent in this challenging context. As a result, the CDR will decrease the negative impacts of physicians' inherent biases and practice disparities, significantly increase overall AHT detection, decrease unnecessary abuse evaluations (and their associated risks), and reduce AHT-associated health care costs—all while lessening physicians' fear of making the "wrong" decision. Most importantly, however, the CDR will save the lives of children who would otherwise fall through the cracks--the head-injured children who are abused but who would be missed using current AHT evaluation criteria. Approximately 35,000 children under the age of 3 present with head trauma injuries admitted to PICUs each year in the U.S.[63, 64] Preliminary studies suggest that although 76% of these presenting cases are high risk (~26,000), only 73% of these cases are currently evaluated for abuse (~19,000). We expect that adopting the CDR will lead to increased evaluation rates of high-risk children by 17%--from 73% to 90%, or an additional 5,000 abuse evaluations. Published estimates indicate that AHT is missed or unrecognized in 30% of head-injured children, more than 25% of children with unrecognized AHT suffer additional inflicted injuries, and 10% of children with unrecognized AHT subsequently die.[8] Broad uptake of our CDR would translate into roughly 1400 cases of AHT being recognized that would have been missed, preventing re-abuse of about 350 children and saving the lives of nearly 100 children *every year*. Note that these are conservative estimates based on a 90% CDR utilization. These numbers would be even more staggering if the CDR were to become the standard of care and if its utilization rates were to rise to 100% nationally. Moreover, having a clear decision rule that is recommended as standard of care would provide physicians and emergency room Providers increased confidence in their decision to evaluate high risk children for abuse, thus reducing bias and alleviating potential fear of litigation for having done so. If the proposed SCRT is successful, the TCCMS will leverage these findings to change practice. For example, a demonstration project for the DOC's Aim 1 is to calculate the hospital 'return-on-investment' (ROI) from CDR adoption. To ensure that this is accomplished, the budget for Project 1 includes paying sites for an additional month after the SCRT is complete to collet the data from for the ROI from the EMR (e.g., length of stay; see Table D1). These potential cost-savings, along with the SCRT findings, will be packaged by the DOC into translational messages designed to resonate with hospital administrators, policy-makers, and practitioners. Finally, the Admin Core Aim 3 includes plans to work with the American Academy of Pediatrics (AAP) toward a formal endorsement of the CDR as standard of care (see letter of support from Dr. Robert Block; former AAP President).