METHODOLOGIC ISSUES

Evaluation of a systematic approach for identifying injury scenarios

Katherine Kaufer Christoffel, Joseph L Schofer, and the Kids’n’Cars Team*

Abstract

Objective—To assess the effectiveness of a new multidisciplinary method for reconstructing the causal sequences that lead to child pedestrian injuries.

Setting—Subjects were 5–12 year old residents of Chicago, Illinois, USA, presenting for care due to pedestrian injury at one pediatric trauma center.

Methods—The interactions of medical, child, psychosocial, and traffic factors contributing to the injury were analysed. For 142 cases, information about the victim, his/her family, the injury site, and the activities just before the injury, was used in a structured manner by a multidisciplinary team to produce injury scenarios. Each scenario comprised a list of contributing factors, an estimate of the importance of each, and a narrative description of the causal sequence leading to the injury event. Face validity was assessed by two outside teams that performed a structured review of a subsample of cases (n = 11). Reliability was evaluated by comparison of the results of parallel teams assessing the same cases (n = 14). Process consistency and bias were assessed by analysis of the correlations of factor-importance rating patterns between members and over time.

Results—The outside team’s agreement scores were based on a 1–5 Likert scale; these showed a mean of 3.6 and median of 4.0. Parallel teams consistently showed agreement greater than 85% on global attributes of cases. Intraclass correlation coefficient scores showed fair or better agreement for all classes of contributors, and excellent agreement for more than one third. Rating pattern analyses showed strong agreement by team members. Agreement did not increase over the period of the study.

Conclusions—This causal sequence reconstruction method has acceptable face validity, reliability, and internal consistency. Although labor intensive and thus costly, it can produce unique, rich information for understanding injury causation and for guiding the search for promising interventions.


Keywords: pedestrian, causal sequence, evaluation, safety.

*The Kids’n’Cars Team members are listed at the end of the paper.

The use of multidisciplinary teams to investigate motor vehicle crashes is not new. Some of the earliest work was done by Baker and others at the Traffic Institute of Northwestern University in the late 1950s. Their team included an engineer, psychologist, and physician, who performed ‘hot pursuit’ investigations of urban auto crashes. The team received notification of events from police and traveled to the scene to examine the site and vehicles, and to interview drivers, victims, and witnesses. Some interviews were conducted at the scene and others in the study office within a few days of the event. Only a small number of crashes was investigated, and most of these were vehicle-vehicle events. A brief description of the event was produced based on group consensus. These researchers were not able to reach consensus on causal factors and their importance. Instead, each disciplinary specialist prepared a separate, descriptive report identifying factors but not assessing their importance.

Snyder and Knoblauch also conducted a 12-city investigation of 2157 pedestrian injury events involving persons of all ages; 40% of the victims were under 9 years of age. Each case was investigated by one or two specialists trained to use a consistent data collection protocol created by a multidisciplinary team. Victims, drivers, and witnesses were interviewed to develop a description of the ‘behavioural sequence’ leading to the injury. About 16% of the cases were investigated at the scene within two hours of the event and nearly 35% were investigated one or more weeks later. Data were not collected on the detailed characteristics of injured pedestrians (beyond age and gender), although driver characteristics were analyzed. No formal process was reported for integrating information into injury scenarios, though cases were classified into 26 scenario categories.

The work described here builds on these earlier efforts. When we began to study child pedestrian injury, we were impressed by the complexity of the causal processes and frustrated by the absence of an appropriate method to weave them into comprehensive scenarios. We sought to identify the psychosocial and physical contributors to the event, to assess their relative importance, and to understand their interactions. This demanded not only a systematic approach to data collection, but also an assessment of these data. We designed a team process that enabled us to integrate several sources of data: descriptions of the event, the scene, and the child’s physical,
social, and psychological characteristics. This process, which we refer to as causal sequence reconstruction (CSR), differs from previous efforts principally because of this multidisciplinary integration and interpretation phase. This report describes an evaluation of the CSR method.

Methods
The methods of the Kids' n' Cars study have been described, and are summarized briefly here. Pedestrian injury victims aged 5–12 years were identified at Children's Memorial Hospital, a pediatric trauma center on the north side of Chicago. Data on each child and injury event were gathered from parents, victims (where feasible), teachers, on-scene investigations, and police accident reports, which often included statements from drivers, witnesses, and victims.

The injury site was visited, measured, videotaped, and diagrammed, and relevant traffic measures were collected at the same time of day and day of week as the event. Each child was characterized by his or her physical attributes and limitations, medical history, injuries, and psychological profile measured with standardized test scores. The Achenbach child behavior checklist was used to evaluate behavior, and the Alpern-Boll developmental profile to assess developmental level. The child's family environment was described by scores on the Moos family environment scale, the Coddington life stress scale, the Thurstone temperament schedule (for mothers), and the multidimensional scales of perceived social support (also about the mother). To minimize the effect of the injury on parental reports of behavior, all measures were obtained as soon as possible after injury.

TEAM PROCESS
Our multidisciplinary team included experts from traffic engineering, medicine, clinical psychology, and social work. Each case was investigated by each of these disciplines following an established and consistent protocol. The findings of each expert were integrated at a group meeting during which each field investigator presented his or her findings in objective terms, for example, interview content, test scores, site measurements, and videotape. This was followed by an open discussion in which a description of the causal sequence was formulated and contributing factors identified from a list of 75 factors in 11 realms, developed interactively early in our study (see Appendix for factors and table 2 for realms). The importance of these factors in contributing to the injury event was determined by an open rating process in which each team member assigned weights on a scale of 0 (not a contributor) to 5 (decisive factor). This process fostered debate and discussion, leading to selection of consensus weights, the modal value of the group ratings. A qualitative (yes/no) judgment was made as to whether the child moved quickly or appeared suddenly (sudden appearance pedestrian injury, SAPI). Finally, a narrative description of each case was prepared by the discussion leader and ratified by the team, which then voted its confidence in the resulting narrative. The mean and mode confidence score was 4 on a scale of 1–5. Meetings took 35–75 minutes; cases were reassessed if new information became available, for example, if a teacher's child behavior checklist arrived belatedly.

EVALUATION
Efforts to assess the soundness of the CSR process and its product scenarios included internal and external reliability, face validity, and analysis of internal consistency and biases (figure).

Internal reliability of the process was assessed by applying it through two parallel teams formed by dividing the Kids'n'Cars (henceforth 'Kids') team into two — Kids A and Kids B — with each discipline represented on each team. The A and B teams were created by random assignment of each discipline's members (usually two) and of team members not tied to a discipline (for example, project coordinator), with new assignments after every two split cases. Six cases were analyzed in this way.

The second component of the evaluation was an assessment of external reliability for eight cases by an outside consultant team (team 1); a two member team from another institution that reviewed the scenarios produced by the Kids team. Team 1 was provided with the same information as the Kids team, that is, 1–2 page summaries of the medical records review, social work interview, psychological scale scores, and traffic site investigations, along with a videotape of the injury site. This allowed us to compare the Kids team results with team 1's results as if they were split teams.

The third component was an assessment of face validity by team 1 and another, team 2. Team 2 was also a two person team, from a third institution. It received the same kind of information as team 1 received for assessment of external reliability. These teams were also given summaries of the Kids team results for each case, including the factors selected as operative, the importance ratings, the SAPI and confidence judgments, and the narrative descriptions. For logistical reasons, one consulting team reviewed eight cases and the other only five; two cases were reviewed by both teams. The two outside teams were asked to review the Kids team's CSR process for the cases they received. Based on those reviews, they were asked to assess, for each of the four data types used — medical, social work, psychological, and traffic — data usefulness, data comprehensiveness, the Kids team's confidence in the data, the contributors the Kids team identified and the weights it assigned to these. The consultants were also asked to rate the soundness of the Kids team's narrative description and the confidence ratings assigned to these. Teams 1 and 2 reported their degree of agreement with the Kids team results, using
five point Likert scales for all statements asserting a positive view of the items being evaluated.

The fourth evaluation component assessed internal consistency and bias by examining the correlations among Kids team members' rating patterns for three realms — road hazards, child distraction, and supervision. The aim was to evaluate (a) whether team leaders exerted undue influence on the rating patterns, and (b) whether the correlations increased over the four year period of the CSR meetings (which would suggest a group dynamic affecting scenario content).

**DATA ANALYSIS**

Outside team evaluations of the Kids team assessments were described using mean and modal Likert scale scores for each component. The two types of two team data (Kids A vs. Kids B and Kids v team 1) were analyzed in two ways. First, the per cent agreement was calculated on several types of team ratings: identification of sudden appearance events, narrative confidence score, and contributor selection and weights. These analyses used the 11 contributor realms listed in table 2. For each realm, the maximum score for any selected contributor for each case was the indicator of the realm's contribution to the causal sequence. The agreement measures used were (a) per cent agreement on selected (and unselected) individual contributors; (b) per cent agreement on selected, that is, realms with any contributor identified and unselected; (c) per cent agreement on maximum contributor weights in selected realms; and (d) disagreement on maximum contributor weights, as indicated by the mean square difference in weights for each realm.

For the Kids A vs Kids B and Kids v team 1 analyses, we also calculated three variance components: team, case, and residual; this was based on a random effects analysis of variance. For each category, the intraclass correlation coefficient (ICC), equal to the case variance component divided by the sum of the three variance components, was calculated. (This statistic is equivalent to a weighted k analysis with weights equal to the square of the difference between scores.) Key statistics include the size of the ICC and the size of the random variance component. We evaluated ICCs as follows: $\geq 0.80$, excellent agreement; $0.60 - 0.80$, good agreement; $0.40 - 0.60$, moderate agreement; $0.20 - 0.40$, fair agreement; and $\leq 0.20$, poor agreement.\(^{15}\)

Internal consistency and bias were analyzed by obtaining correlations for every combination of two voters for each realm analyzed. Each voter was then characterized by the mean correlation with other voters. In addition, cases were divided into four equal sized groups in four successive time periods. A general linear models procedure was used to generate an intraclass correlation coefficient and mean square error for votes in each time period. These results were examined for evidence of a temporal learning effect, as evidenced by a decreasing mean square error and increasing ICC.

**Results**

**FACE VALIDITY**

As described under the methods section face validity was assessed using a five point Likert scale (1 = strongly disagree to 5 = strongly agree). The mean scores for 24 items, for each outside team, and an overall mean, across cases.
and teams were ≥ 3·5 for 21 items and > 4 for 11 items. For the cases staffed by both outside teams, the mean was 3·6, and the median 4·0.

TWO TEAM DATA
Table 1 shows that there was relatively strong agreement between parallel teams' results. Agreement was strongest for the six Kids A v Kids B comparison and was not substantially weaker for the eight Kids v team 1 cases. Table 2 shows the variance analyses for the Kids A v Kids B and Kids v team 1 comparisons. The ICC values were above 0·80 for six of the 13 Kids v Kids comparisons. One comparison showed an ICC of 0; this was due to lack of subject variability: the realm was selected in only two cases and scored as low in importance in both. ICC values were also above 0·80 for five of the Kids v team 1 comparisons and were 0·20 or better for all remaining comparisons. This table shows that both kinds of split teams functioned equally well.

In the comparisons in both tables 1 and 2, the teams agreed best on the most global measures (for example, sudden appearance), and least on the most detailed assessments, that it, maximum factor weightings within realms.

RATING PATTERNS
Because of the large number of team members and contributing factors rated, the inter-rater analyses generated numerous correlation values. Table 3 shows selected values that exemplify these analyses. For each team member listed, the mean and range of the correlations with all other team members is shown only for the road characteristics factor realm. Correlations were high and did not differ much among team members nor were they higher for the senior team members than for others. The other results analyzed were similar.

The ICC and mean square errors for votes on road hazards in successive time periods varied from 0·92 in the first period to 0·76 in the third (mean square error varied from 0·303 to 0·706). The differences were not statistically significant. Thus the pattern observed does not support an inference of tightened team dynamics (and so possibly bias) over time.

Discussion
Preventing injury requires an understanding of the factors and events leading to the tissue damaging energy exchange. Only with this understanding is it possible to identify effective means to interrupt the process. Examples of this principle include the development of motor vehicle occupant restraint systems to safely displace crash forces, the identification of four sided fences with self locking gates as a means to delay pool access long enough to prevent toddler drownings, and the development of 'Wilky Whistle', an approach to teaching children to cross streets safely.6 7

One approach commonly used to reach this understanding is an examination of the sequence of events to detect patterns that suggest causal factors and prevention opportunities. This process generates scenarios that allow a qualitative understanding of how injury events happen; it therefore enables investigators to set priorities for data collection, data analysis, and the design and evaluation of preventive interventions. The scenarios are also useful for communicating with other investigators and injury control practitioners. Yet scenario construction has been a hidden step in injury prevention work.

Despite widespread use of this general approach, consistent, verifiable procedures are rarely used and the process may therefore be seen as unscientific. The utility of prior multidisciplinary efforts is based on a presumption of reliability and validity rather than any sorts of quantitative evaluation. Current standards require rigorous scrutiny of complex, expensive, and/or novel methods. Evaluation is par-
particularly needed for the types of process we describe because of the qualitative nature of the resulting scenarios, which is the source of both their value and their uncertainty. If any CSR method does not produce consistent, reliable, and unbiased results, it is merely an anecdotal process without credibility. If this CSR process can be structured with rigor and demonstrated to be valid, the importance of scenario constructions could be more fully acknowledged, its products more credible, and its results more widely shared.

The CSR method described here responds to a problem faced by all who study rare, random events: almost never do investigators witness the event or the steps leading to it. Instead, researchers must rely on witness reports that are deficient in many ways. Nor do trained professionals (for example, police officers) often witness these events directly. Our work did not involve immediate visits to the injury site because of the costs and risks involved. Instead, we reconstructed each event using information from several different sources. These include parent descriptions that often were secondhand (from victims, friends, and siblings). When possible, however, we also interviewed the injured child. Police reports implicitly or explicitly include the driver’s explanation, and sometimes witness descriptions. Some or all of these observers and reporters may distort the facts, unconsciously (for example, to provide a rationale for a horrible event) or consciously (for example, to avoid or assign blame, cover up a misdemeanor or rule infraction). In such cases it is not possible to establish a ‘gold standard’ measurement of what really happened.

The CSR process described here integrates information from several sources through a logical, open replicable procedure that produces ‘most likely’ process description. At the core of this method is the team meeting — an interactive process for challenging and cross checking reports from various data sources to reduce bias and increase objectivity.

The nature of this approach demands evaluation to assess (1) internal and external reliability, accomplished here by two team analyses; (2) face validity, here done through evaluation by consultant teams; and (3) internal consistency and bias, pursued here through an assessment of the contribution of group dynamics to process outcomes. The number of cases studied by consultant teams included about 10% of the sample of cases for which scenarios were constructed.

INTERNAL AND EXTERNAL RELIABILITY (TWO TEAM ANALYSES)

In view of the structural uncertainties built into the CSR method, the level of agreement shown in the two team analyses is remarkably high, with substantial difficulties only on a few specific contributor weightings. Reliability was extremely good for the least detailed analyses (for example, confidence in narrative and sudden appearance), and excellent on more than one third of the contributing realms.

Agreement was moderate to good on the role of the driver, although we expected it to be poor because of the uncertainty inherent in this area. Driver behavior is typically recorded in police reports, with drivers themselves as the informant. Driver bias and social expectations lead to a tendency to blame the child. Our approach was more comprehensive, combining reports from police, the child, parents, and other witnesses.

Agreement was also surprisingly good for newly identified contributors, such as sudden appearance. For these, such agreement might have resulted from shared experience, which the Kids split team members had in greater abundance than the team 1 members. Excellent agreement across all teams on these is a reassuring indication of the utility of these newly identified contributors.

Reliability was weaker for several contributor realms. In the case of traffic contributors, which seem objective, this result is counterintuitive. However, as the injury event was not witnessed, the contribution of particular traffic configurations is somewhat speculative. Transient conditions are particularly susceptible to variable reporting from different sources. Thus, our modest reliability results on traffic contributors are not surprising. What is most important is that agreement on the injury narrative was not adversely affected by modest agreement concerning traffic contributors.

For a few contributor realms, agreement levels were notably discrepant between the two sets of two team comparisons (that is, Kids A v Kids B, Kids v team 1). Lack of agreement on ‘medical’ contributors could easily have been caused by their rarity. Further, the role of medical problems was not always clear, and the team 1 physician may not have seen this quite as clearly as the three medical people on the Kids team (two physicians (MDs) and one registered nurse), who discussed many more cases and became used to the fact that medical issues only rarely contributed to the causal sequence.

The Kids v team 1 comparison for ‘child inner processes’ was moderate, but substantially lower than the excellent agreement for the Kids A v Kids B comparison. This may have reflected some disagreement between the Kids and team 1 psychologists about the interpretation of subscale scores from the Achenbach child behaviour checklist. In addition, the precise connection between child characteristics and unobserved behaviors during the causal sequence is, like the role of traffic, subject to speculation. Again, what is most important is that the existing level of uncertainty did not undermine agreement on the injury narratives.

The discrepancy on the view/visibility contributors, with very good reliability for the Kids v team 1 comparison but only fair agreement for the Kids A v Kids B comparison, is perplexing, particularly given the high interteam consistency on ‘sudden appearance’.

FACE VALIDITY

The evaluations indicate that this CSR method appears reasonably sound to outside reviewers.
INTERNAL CONSISTENCY AND BIAS (RATING PATTERNS)
This analysis was conducted for the 120 cases with complete records of the relevant ratings. Intervenor agreement was high, but not perfect. Clearly, the open voting process did not suppress all disagreement. We were concerned that it might have given undue influence to team leaders but this did not occur. The analyses of sequential subgroups of cases also provide assurance that team interactions did not lead to increasingly idiosyncratic results over time.

SIGNIFICANCE OF EVALUATION RESULTS
The specific type of data used differed somewhat from previous CSR efforts: our study collected more psychosocial information than earlier efforts but did not interview drivers or examine vehicles. Although driver reports can be expected to carry substantial biases, the absence of all but the simplest objective information about drivers from the police reports represents a gap in our data. However, this CSR method can readily accommodate data from various sources.

APPLICATION OF THE METHOD
Our CSR method produced reliable scenarios, with face validity, that were not shaped by idiosyncratic group dynamics. This method may be useful to other investigators studying other complex types of injury. The resulting scenarios appear to be credible process descriptions, and thus components of future injury research. Thus, this method appears to be effective for understanding complex interactions of qualitative contributors. These include definitional, measurement, and conceptualization issues which may be at the core of understanding injury event causality.

For example, while it is intuitive that supervision of children is important in, and protective against, pedestrain injury, early in our work we found that children were getting hit by cars even when supervised by parents. Further discussions made it clear that there was no operational definition of supervision to help us understand such cases. Our method led us to a four dimensional definition of supervision: supervised or not, age of the supervisor, distance of the supervisor from the child, and whether the child is alone or with a group of peers.1

Similarly, when we tried to fit our cases into the accepted pedestrain injury taxonomy,2 we found ambiguities that led us to separate so-called dart-out events into two distinct but often associated phenomena: moving quickly into traffic (running across the street or changing directions rapidly while in the street); and appearing suddenly, not necessarily due to rapid movement, but because of view obstructions from parked cars, cars stopped in traffic, or street furniture.3 This distinction clarifies the process of the injury event and may lead to separate interventions designed to slow children down and make them more visible.

This CSR approach is particularly appropri-
Child pedestrian injury scenarios

Example partial text:

- Trait discrepancy, for example, physical less than social
- Internalizing psychiatric disturbance
- Wide street
- Normally high speed street
- Complex street configuration
- Complex operations
- Environmental impediment to pedestrian traffic makes pedestrian unexpected
- Wet pavement
- High traffic volume
- Light traffic volume (deceptively safe)
- Distracted by events/conditions inside vehicle
- Distracted by events/conditions outside vehicle
- Driving too fast for conditions
- Other dangerous/reckless driver action
- Child exited car, entering traffic inappropriately
- View (by or of child) obstructed by fixed objects
- View (by or of child) obstructed by standing or moving vehicles
- View (by or of child) obstructed by legally parked vehicles
- View (by or of child) obstructed by double parked or illegally parked vehicles
- View (by or of child) restricted because of child's height
- Low visibility (for driver, child) due to weather
- Low visibility (for driver, child) due to lighting (night)
- Child negotiated part-way across street and was hit in second half of street (continued across)
- Child negotiated part-way across street (and was hit) returning to safety of original side
- Child negotiated part-way across street (and was hit) facilitated by encouragement by driver(s)
- Child attracted/distracted by people at site
- Child attracted/distracted by place or object (not in street)
- Child followed peers into street ("trailer")
- Child followed adult into street
- Child playing, game spilled into street
- Child attracted by area conducive to play
- Child escaping a situation by going into street or being chased
- Child attracted by complexity of visual environment
- Child distracted/confused by high noise levels at site
- Pressure specific: 'hurry and cross'
- Group crossing
- Special event
- Time pressed
- Child attempted midblock crossing for convenience, 'propinquity' of attraction
- Child attempted midblock crossing — appeared less complex than intersection option
- Child attempted midblock crossing — was environmentally directed (design promotes midblock crossing)
- Moved quickly
- Shoe trouble
- Entered street to retrieve object
- Skateboard or other object used as scooter
- Child's experience with this site: unfamiliar, not experienced
- Child's experience with this site: very familiar, careless
- Intersection crossing, avoiding more common site
- Child's experience with site of this type: unfamiliar, not experienced
- Child's experience with site of this type: very familiar, careless
- Lack of appropriate control
- Crosswalk (deceptively safe)

References:

4. Distracted by View dangerous/reckless Child Child visibility (for driver, child)
11. Baumrind D. Child care practices antecedent to three patterns of preschool behavior. Gen Psych Monographs 1967; 75: