

1-1-2015

Accuracy of physician-estimated probability of brain injury in children with minor head trauma

Carrie Daymont
University of Manitoba

Terry P. Klassen
University of Manitoba

Martin H. Osmond
University of Ottawa

Keith Aronyk
University of Alberta

Benoit Bailey
CHU Sainte-Justine - Le Centre Hospitalier Universitaire Mère-Enfant

See next page for additional authors

Follow this and additional works at: <https://ir.lib.uwo.ca/paedpub>

Citation of this paper:

Daymont, Carrie; Klassen, Terry P.; Osmond, Martin H.; Aronyk, Keith; Bailey, Benoit; Chauvin-Kimoff, Laurel; Hamilton, Mark; and Anna Jarvis, D., "Accuracy of physician-estimated probability of brain injury in children with minor head trauma" (2015). *Paediatrics Publications*. 1514.
<https://ir.lib.uwo.ca/paedpub/1514>

Authors

Carrie Daymont, Terry P. Klassen, Martin H. Osmond, Keith Aronyk, Benoit Bailey, Laurel Chauvin-Kimoff, Mark Hamilton, and D. Anna Jarvis

Accuracy of physician-estimated probability of brain injury in children with minor head trauma

Carrie Daymont, MD, MSCE*^{†‡}; Terry P. Klassen, MD, MSc*^{†‡}; Martin H. Osmond, MD, CM^{§¶}
(for the Pediatric Emergency Research Canada [PERC] Head Injury Study Group)

ABSTRACT

Objective: To evaluate the accuracy of physician estimates of the probability of intracranial injury in children with minor head trauma.

Methods: This is a subanalysis of a large prospective multicentre cohort study performed from July 2001 to November 2005. During data collection for the derivation of a clinical prediction rule for children with minor head trauma, physicians indicated their estimate of the probability of brain injury visible on computed tomography (P-Injury) and the probability of injury requiring intervention (P-Intervention) by choosing one of the following options: 0%, 1%, 2%, 3%, 4%, 5%, 10%, 20%, 30%, 40%, 50%, 75%, 90%, and 100%. We compared observed frequencies to expected frequencies of injury using Pearson's χ^2 -test in analyses stratified by the level of each type of predicted probability and by year of age.

Results: In 3771 eligible subjects, the mean predicted risk was 4.6% (P-Injury) and 1.4% (P-Intervention). The observed frequency of injury was 4.1% (any injury) and 0.6% (intervention). For all levels of P-Injury from 1% to 40%, the observed frequency of injury was consistent with the expected frequency. The observed frequencies for the 50%, 75%, and 90% levels were lower than expected ($p < 0.05$). For estimates of P-Intervention, the observed frequency was consistently higher than the expected frequency. Physicians underestimated risk for infants (mean P-Intervention 6.2%, actual risk 12.3%, $p < 0.001$).

Conclusions: Physician estimates of probability of any brain injury in children were collectively accurate for children with low and moderate degrees of predicted risk. Risk was underestimated in infants.

RÉSUMÉ

Objectif: L'étude visait à évaluer l'exactitude de l'estimation, faite par les médecins, du risque de lésion intracrânienne chez les enfants ayant subi un trauma léger à la tête.

Méthode: Il s'agit d'une analyse secondaire d'une importante étude de cohortes, multicentrique et prospective, menée de juillet 2001 à novembre 2005. Au cours du processus de

collecte de données en vue de dégager une règle de prévision clinique chez les enfants ayant subi un trauma léger à la tête, les médecins ont indiqué leur estimation du risque de lésion cérébrale visible à la tomodensitométrie (probabilité de lésion [P lésion]) et du risque de lésion nécessitant une intervention (probabilité d'intervention [P intervention]) en indiquant l'une des valeurs suivantes: 0 %, 1 %, 2 %, 3 %, 4 %, 5 %, 10 %, 20 %, 30 %, 40 %, 50 %, 75 %, 90 % ou 100 %. Les auteurs ont comparé les fréquences observées avec les fréquences prévues de lésion à l'aide du test χ^2 de Pearson dans des analyses divisées selon le degré de chaque type de probabilité prévue et selon l'âge.

Résultats: Le risque moyen prévu chez 3771 sujets admissibles était de 4,6 % (P lésion) et de 1,4 % (P intervention). Quant à la fréquence observée des lésions, elle s'élevait à 4,1 % pour l'ensemble des lésions et à 0,6 % pour les interventions. En ce qui concerne la fréquence observée des lésions, elle correspondait à la fréquence prévue, et ce, à toutes les valeurs variant de 1 % à 40 %; pour ce qui est de la fréquence observée aux valeurs de 50 %, 75 % et 90 %, elle était inférieure à la fréquence prévue ($p < 0,05$). En ce qui concerne l'estimation des probabilités d'intervention, la fréquence observée était toujours plus élevée que la fréquence prévue. Enfin, les médecins sous-estimaient le risque chez les nourrissons (P intervention moyenne: 6,2 %; risque réel: 12,3 %; $p < 0,001$).

Conclusions: L'estimation, faite par les médecins, du risque de lésion cérébrale chez les enfants était exacte, dans l'ensemble, chez les blessés ayant un risque prévu faible ou moyen; par contre, le risque était sous-estimé chez les nourrissons.

Keywords: head injury, traumatic brain injury, pediatrics, risk, decision-making

INTRODUCTION

Many clinical decisions involve an evaluation and comparison of risks and benefits. Little is known about how physicians estimate risks and benefits or how those

From the *Department of Pediatrics and Child Health, University of Manitoba, Winnipeg, MB; †Children's Hospital Research Institute of Manitoba, Winnipeg, MB; ‡The George and Fay Yee Centre for Healthcare Innovation, University of Manitoba, Winnipeg, MB; §Department of Pediatrics, University of Ottawa, ON; and ¶The Children's Hospital of Eastern Ontario Research Institute, University of Ottawa, ON

Correspondence to: Dr. Carrie Daymont, Children's Hospital Research Institute of Manitoba, 655A-715 McDermot Avenue, Winnipeg, MB R3E 0Z2; Email: cdaymont@chrimg.ca

estimations are used in decision-making. Although many studies have evaluated physician pretest probabilities, relatively few studies have prospectively compared physicians' quantitative estimates of the risk of a condition with the true probability of that condition in real patients.¹⁻⁶ In such studies, most showed that physicians provided accurate probability estimates in one or more subsets of patients or clinical situations but provided inaccurate estimates in other subsets. Prior research has identified wide variability between physicians regarding estimates of the probability of an outcome for the same patient.^{3,7} To our knowledge, estimates of the probability of outcomes have not been evaluated previously for patients with head injury.

An improved understanding of how physicians estimate risk and the degree to which they are accurate may facilitate promotion of effective decision-making strategies and warnings against ineffective decision strategies, both for children with head injury and for other conditions.^{8,9} A greater understanding of the influences that impact physician decision-making may also facilitate the design of clinical practice guidelines that incorporate all information considered relevant by physicians and are more likely to influence physician behaviour.^{10,11}

In order to explore these issues further, we were able to use data obtained during derivation of a clinical prediction rule for predicting brain injury in children with minor head trauma. After assessing patients and before any diagnostic imaging, physicians were asked to provide an estimate of the probability of any brain injury and an injury requiring intervention.¹² Using these data, we sought to determine the accuracy of physicians' probability estimates, to explore influences on probability estimates, and to explore the relationship between probability estimates and decisions.

METHODS

We performed a secondary analysis of data from the Canadian Assessment of Tomography for Childhood Head Injury (CATCH) derivation study, a prospective multicentre observational study used to derive a clinical prediction rule in children with a minor head injury.¹² Subjects were required to have a Glasgow Coma Score on presentation of 13 or higher and at least one of the following: witnessed loss of consciousness or disorientation, definite amnesia, persistent vomiting, or persistent irritability in the emergency department (if <2 years old). During the derivation study,

physicians were asked to report clinical characteristics of included patients at initial assessment and to perform further evaluation as they deemed appropriate. All included subjects who did not undergo computed tomography (CT) during the initial visit and received a phone follow-up 2 weeks after discharge in order to identify any clinically significant injuries that were not previously found. Further details of the methods have been previously reported.¹²

The research ethics boards at the University of Ottawa and the University of Manitoba approved this secondary analysis. Statistical testing was performed using Stata 13.1 (StataCorp LP, College Station, Texas).

Measures

In addition to an assessment of specific clinical characteristics, physicians assessing patients for CATCH were asked to choose a value corresponding to their estimate of the probability of any brain injury (P-Injury), an estimate of the probability of an injury requiring neurologic intervention (P-Intervention), and a rating on a five-point scale of how comfortable they would feel not ordering a CT scan (Figure 1). A notice on the case report form reminded physicians that the questions should be answered before any imaging was done. Brain injury was defined as "any acute intracranial finding revealed on CT that was attributable to acute injury," including depressed skull fractures but excluding nondepressed and basilar skull fractures.¹² The need for neurologic intervention was defined as having a craniotomy, intubation, elevation of skull fracture, intracranial pressure monitoring, or death within 7 days. Study assessments were completed by attending physicians or resident physicians in their second year of training or beyond in pediatrics, emergency medicine, or family medicine.

Accuracy of estimates

We used Pearson's χ^2 -test to compare expected and observed frequencies of subjects with injury for each level of P-Injury and P-Intervention. This test could not be performed for certain levels of risk, including 0% and 100%, for which the expected frequency of subjects with or without injury was <2. We also evaluated the overall goodness-of-fit of P-Injury and P-Intervention using the Hosmer-Lemeshow test.

PREDICT PROBABILITY OF: (Circle response) <u>MUST BE COMPLETED PRIOR TO CT</u>	
<u>Acute brain injury seen on CT scan</u>	
0% 1% 2% 3% 4% 5% 10% 20% 30% 40% 50% 75% 90% 100%	
<u>Patient requiring neurological intervention</u>	
0% 1% 2% 3% 4% 5% 10% 20% 30% 40% 50% 75% 90% 100%	
How comfortable would you be not ordering a CT scan for this patient?	
<input type="checkbox"/> Very comfortable <input type="checkbox"/> Comfortable <input type="checkbox"/> Neutral <input type="checkbox"/> Uncomfortable <input type="checkbox"/> Very Uncomfortable	

Figure 1. Probability estimate and comfort rating items on physician assessment forms.

Patient characteristics

For subjects in each year of age and for each sex, we compared the expected and actual numbers of patients with any brain injury using Pearson's χ^2 -test.

Decisions

We used two measures of physician decision-making that provided complementary information: 1) whether a patient received a CT scan and 2) the rating of how comfortable a physician was with not ordering a CT scan, which was performed at the time of the initial assessment. For some analyses about decisions, we created a binary measure of physician comfort: *comfortable* (i.e., either comfortable or very comfortable not ordering a CT) or *not comfortable* (i.e., neutral, uncomfortable, or very uncomfortable not ordering a CT). These categories were based on the proportion of subjects with each comfort rating who received a CT. For each level of P-Injury and P-Intervention, we determined the proportion of patients who underwent CT and the proportion that received a *not comfortable* rating. We used logistic regression to determine whether there was a difference in the relationship between estimated probability of injury and CT performance in young children.

Inter-rater agreement

A subset of subjects had assessments by two physicians. Both attending and resident physicians could serve as the primary or secondary rater. Inter-rater agreement

was determined for P-Injury, P-Intervention, and comfort ratings using weighted κ statistics (kap command in Stata 13.1 using w2 option for weights). For these evaluations, we used a numerical point scale for comfort ratings (1 = very uncomfortable not ordering a CT, 5 = very comfortable ordering a CT).

Physician employment status

We stratified the sample by the physician employment status of the primary rater (full-time attending, part-time attending, or resident) in order to explore differences in estimates and decisions between the groups. For each group, we used Pearson's χ^2 -test to compare expected and overall frequencies of any brain injury. We also compared the frequency of injury, probability estimates, mean comfort ratings, and the proportion undergoing CT for attending v. resident physicians using the Kruskal-Wallis test for continuous variables and Pearson's χ^2 -test for binary variables.

Test characteristics

We determined the test characteristics (i.e., sensitivity, specificity, positive predictive value, and negative predictive value) of physician-estimated probability using each level of P-Injury and P-Intervention as a cutpoint. We compared the test characteristics of the CATCH clinical prediction rule to the test characteristics of physician-estimated probability for each cutpoint with a sensitivity of at least 95%.

Table 1. Selected demographic and outcome characteristics, presented by age group

	Overall	<1 year	1 to <2 years	2 to < years	6 to <12 years	12 to 17 years
Subjects = <i>n</i>	3771	179	90	761	1240	1501
Male (<i>n</i>)	2439	78	48	417	790	1106
(%)	58%	44%	53%	55%	64%	74%
CT performed (<i>n</i>)	1986	96	34	361	662	833
(%)	53%	54%	38%	47%	53%	56%
Brain injury (<i>n</i>)	153	22	1	25	59	46
(%)	4.1%	12.3%	1.1%	3.3%	4.8%	3.1%
Received neurologic intervention (<i>n</i>)	24	1	0	4	10	9
(%)	0.6%	0.6%	0%	0.5%	0.8%	0.6%

RESULTS

Enrolled subjects ($n = 3866$) were assessed by 1150 physicians at 10 centres (Table 1). All subjects who were ultimately diagnosed with brain injury underwent CT prior to discharge. Further details of subject characteristics have been published previously.¹²

Physician probability estimates were missing for approximately 2% ($n = 74$) of the enrolled subjects. Subjects from 1 of the 10 centres and subjects for whom the primary rater was a resident physician were more likely to have missing probability estimates. Otherwise, no differences between subjects with and without missing estimates were identified. For 21 (0.5%) ratings, P-Intervention was greater than P-Injury. These values were excluded from further analysis because they were internally inconsistent and likely to represent errors. The following analyses were performed on 3771 subjects with consistent non-missing values for P-Injury and P-Intervention, including 153 subjects with brain injury and 24 with brain injury requiring intervention.

Accuracy of estimates

Using a two-tailed χ^2 -test, all levels of P-Injury from 1% to 40% had observed frequencies of injury consistent with the expected frequency (Table 2). The observed frequencies for the 50%, 75%, and 90% levels were lower than expected. For all levels of P-Intervention with an expected frequency of injury of two or greater, the observed frequency of injury was lower than the expected frequency, and this difference was significant for the 1% and 50% levels. Using the Hosmer–Lemeshow test, neither P-Injury ($p = 0.04$) nor P-Intervention ($p < 0.001$) was found to demonstrate overall goodness-of-fit compared to expected frequencies of injury.

Patient characteristics

The observed frequency of injury was consistent with the expected frequency for both sexes. When stratified by age in years of the subject, the observed frequency was consistent with the expected frequency for all ages except zero years ($n = 179$, mean P-Injury 6.2%, actual injury 12.3%, $p = 0.001$) and 12 years ($n = 237$, mean P-Injury 5.6%, actual injury 2.5%, $p = 0.046$).

Decisions

The proportion of subjects receiving a CT scan increased from 9% to 100% with increasing probability estimates (Table 2). The majority of subjects with a P-Injury of 2% or higher and the majority of those with a P-Intervention of 1% or higher received neuroimaging. The proportion of subjects who received a *not comfortable* rating similarly increased, although the proportion receiving a CT scan was generally higher than the proportion with a *not comfortable* rating. Overall, 96% of subjects with a *not comfortable* rating received a CT scan (very uncomfortable 97%, uncomfortable 99%, neutral 90%), and 20% of subjects with a *comfortable* rating received a CT scan (comfortable 40%, very comfortable 8%).

Of the subjects with an initial *comfortable* rating who ultimately received a CT scan, 13 of 426 (3%) were diagnosed with actual injury, and 2 of 426 (0.5%) received intervention. Of the subjects with a *not comfortable* rating who did not receive a CT scan, 0 of 66 were diagnosed with brain injury after discharge.

The relationship between estimated probability of injury and CT performance was not modified by young age (defined either as age less than 1 year or less than 2 years).

Table 2. Proportion of subjects with actual injury and neuroimaging decisions by level of probability estimate

	Subjects (n)	Actual brain injury (n)	Actual brain injury (%; 95% confidence interval)	Injury requiring intervention (n)	Injury requiring intervention (%; 95% confidence interval)	P-value*	Received CT	Not comfortable rating
P-Injury level						0.04		
0%	809	2	0.2 (0.03, 0.9)	0	0 (0, 0.5)	-	9%	3%
1%	953	9	0.9 (0.4, 1.8)	1	0.1 (<0.01, 0.6)	0.8	36%	22%
2%	499	11	2.2 (1.1, 3.9)	1	0.2 (<0.01, 1.1)	0.7	60%	48%
3%	265	7	2.6 (1.1, 5.4)	1	0.4 (0.01, 2.1)	0.7	69%	59%
4%	103	6	5.8 (2.2, 12.2)	1	1.0 (0.02, 5.3)	0.3	74%	73%
5%	667	37	5.6 (3.9, 7.6)	5	0.8 (0.2, 1.7)	0.5	85%	74%
10%	264	26	9.8 (6.5, 14.1)	2	0.8 (0.1, 2.7)	1	94%	90%
20%	78	15	19.2 (11.2, 29.7)	2	2.6 (0.3, 9.0)	0.8	94%	88%
30%	43	12	27.9 (15.3, 43.7)	4	9.3 (2.6, 22.1)	0.7	95%	98%
40%	16	3	18.8 (4.0, 45.6)	1	6.3 (0.2, 30.2)	0.1	100%	100%
50%	43	10	23.3 (11.8, 38.6)	2	4.7 (0.6, 15.8)	<0.001	100%	95%
75%	19	6	31.6 (12.6, 56.6)	1	5.3 (0.1, 26.0)	<0.001	100%	100%
90%	7	4	57.1 (18.4, 90.1)	1	14.3 (0.4, 57.9)	0.03	100%	100%
100%	5	5	100 (47.8, 100)	2	40.0 (5.2, 85.3)	-	100%	100%
P-Intervention level						<0.001		
0%	1881	24	1.3 (0.8, 1.9)	1	0.1 (<0.01, 0.3)	-	31%	20%
1%	1181	41	3.5 (2.5, 4.7)	4	0.3 (0.09, 0.9)	0.02	66%	56%
2%	264	19	7.2 (4.4, 11.0)	2	0.8 (0.09, 2.7)	0.2	83%	78%
3%	94	11	11.7 (6.0, 20.0)	1	1.1 (0.03, 5.8)	0.2	91%	92%
4%	24	1	4.2 (0.1, 21.1)	0	0 (0, 14.2)	-	79%	83%
5%	209	26	12.4 (8.3, 17.7)	5	2.4 (0.8, 5.5)	0.1	87%	82%
10%	70	18	25.7 (16.0, 37.6)	5	7.1 (2.4, 15.9)	0.4	94%	97%
20%	19	2	10.5 (1.3, 33.1)	1	5.3 (0.1, 26.0)	0.09	95%	100%
30%	12	4	33.3 (9.9, 65.1)	2	16.7 (2.1, 48.4)	0.2	100%	100%
40%	2	0	0 (0, 8.42)	0	0 (0, 84.2)	-	100%	100%
50%	14	6	42.9 (17.7, 71.1)	2	14.3 (1.8, 42.8)	0.008	100%	86%
75%	0	-	-	-	-	-	-	-
90%	1	1	100 (2.5, 100)	1	100 (2.5, 100)	-	100%	100%
100%	0	-	-	-	-	-	-	-
Overall	3771	153	4.1 (3.5, 4.7)	24	0.6 (0.4, 0.9)		53%	43%

*P-values for P-Injury and P-Intervention overall are from the Hosmer-Lemeshow test. P-values for individual levels of predicted probabilities are from Pearson's χ^2 -tests with 1 degree of freedom, performed for levels in which all expected frequencies were ≥ 2 .

Inter-rater agreement

There were 333 subjects with two raters; 323 had P-Injury estimates by both raters. Of these, 14 subjects had actual brain injury, and none required intervention. Inter-rater agreement was fair to moderate for P-Injury ($\kappa = 0.46$), P-Intervention ($\kappa = 0.34$), and comfort ratings ($\kappa = 0.60$).

Physician employment status

The majority of patients had a primary rater who was a full-time attending physician (Table 3). For all three

types of physicians, the observed frequency of injury was consistent with the expected frequency. Subjects seen by resident physicians did have significantly higher probability estimates ($p < 0.05$), lower mean comfort ratings, and higher likelihood of CT performance. The difference in the risk of injury by physician type (attending v. resident) was not statistically significant.

Test characteristics

The only cut-offs identified with a sensitivity of at least 95% were a risk level of $>0\%$ for both P-Injury and P-Intervention. These cut-offs generally performed less

Table 3. Stratification by physician employment status

	Full-time attending	Part-time attending	Resident
Subjects (<i>n</i>)	2638	535	598
(%)	70	14	16
Received CT (<i>n</i>)	1350	278	358
(%)	51	52	60
Actual brain injury (<i>n</i>)	103	23	27
(%, 95% Confidence interval)	3.9 (3.2, 4.7)	4.3 (2.7, 6.4)	4.5 (3.0, 6.5)
Mean P-Injury	4.3%	4.3%	6.3%
Mean comfort rating	3.6	3.5	3.4
P-value*	0.4	1	0.07

*P-value from Pearson's χ^2 -test comparing expected with actual rates of any brain injury.

well than the corresponding CATCH rule (P-Injury compared to the medium-risk rule, P-Intervention compared to the high-risk rule).

DISCUSSION

Collectively, physicians in this study accurately estimated the probability of brain injury visible on CT in children with minor head trauma with low and moderate degrees of predicted risk (<50%), which constituted the majority of subjects. Physicians overestimated risk when predicting very high levels of risk of injury or when predicting the probability of injury requiring intervention. Inter-rater agreement for probability estimates was only fair to moderate. Although collective risk estimates for low and moderate degrees of risk were accurate, risk was underestimated in certain populations and overestimated in others. Most notably, risk was underestimated in infants, and the mean estimated probability of injury in children less than 1 year of age was approximately half the true risk of injury. Most children with an estimated probability of any injury of 2% or higher, or an estimated probability of injury of 1% or higher, received a CT scan.

Our results support the continued use of clinical prediction rules for identifying children with head injury who need neuroimaging, even in groups for which risk was predicted accurately. We evaluated predicted and observed risk across a broad range, but the decision to obtain a CT is binary. Clinical prediction rules are designed to maximize discriminative ability in addition to accuracy, and the CATCH rule demonstrated substantially better test characteristics than predicted probability by physicians. Furthermore, physicians in our study working in a pediatric emergency department may be more accurate at predicting risk than physicians who

see children less often. In addition, although estimated risk was collectively accurate, inter-rater agreement was only fair to moderate, indicating that some physicians may have been less accurate than others when estimating risk.

Increasing estimates of the probability of injury were associated with increasing CT use and an increasing frequency of physicians indicating discomfort not ordering a CT. However, within the same level of P-Injury and P-Intervention, some patients underwent CT and some did not. Some variability in the receipt of CT would be expected because of a change in status of the patient. For example, a child who was initially well-appearing may have become lethargic, or a toddler who was initially irritable may have become playful. However, physicians also varied within levels of P-Injury regarding their degree of comfort not ordering a CT scan. This rating was done at the same time as P-Injury and would not have been affected by a later change in the patient's status. This variability of decision-making for patients deemed to have the same probability of a poor outcome is interesting to consider. When clinical decision rules are used, the probability of the outcome is often the only factor that determines the recommended action.

Some of the observed variability in decision-making within levels of P-Injury and P-Intervention are likely secondary to differences between and within individual physicians regarding whether they were seeking to identify all visible injuries or only those injuries that required intervention.^{12,13}

Variability could also be explained by factors beyond the probability of injury that affect the balance of risks and benefits for a specific patient. For example, recent studies have highlighted the increased risk of harm for younger patients from ionizing radiation used in CT.^{14,15}

Table 4. Test characteristics of predicted probabilities compared to the CATCH prediction rule

	P-Injury > 0%	CATCH medium risk* (%)	P-Intervention > 0%	CATCH high risk ¹² (%)
Sensitivity	98.7	98.1	95.8	100
Specificity	22.3	50.1	50.2	70.2
Positive predictive value	5.1	7.8	1.2	2.1
Negative predictive value	99.8	99.8	99.9	100

We did not find effect modification by age on the relationship between estimated probability and decisions, but physicians did underestimate the risk of pathology in children less than 1 year. This underestimation of risk likely reflects the difficulty inherent in assessing infants. However, it is also possible that increased concerns about harm from radiation in very young patients and discomfort about ordering a CT scan somehow made physicians more likely to underestimate the probability of injury in these patients. Variability in decision-making may also be due to varying levels of concern regarding the caregiver's ability to detect clinical worsening at home and obtain prompt follow-up.

Physician decision-making may also have been impacted by factors unrelated to the risks and benefits to the patient. For example, physicians may have been incorporating patient or family preferences into the decision. Features of the emergency department, such as the degree of crowding or how difficult it was to obtain a CT scan, may also have influenced decision-making. Unfortunately, we were unable to assess these potential confounding variables.

Overall, inter-rater reliability of probability estimates was only fair to moderate, confirming previous findings that pretest probabilities for the same patients vary between physicians.^{3,7} Resident physicians generally provided higher estimates of probability of injury, but their mean estimates were consistent with the observed frequency of injury for the children that they rated.

The strengths of our study include the relatively large sample size, the prospective nature of data collection using real patients, the availability of outcome data, including phone follow-up for children who did not undergo CT, and the use of two raters for a subset of subjects.

There are important limitations to our study. The wording of the questions and the choices of risk categories were selected by investigators and have not been validated. Our evaluation of the relationship of probability estimates and CT ordering was limited by the fact that final decisions about CT ordering may not have been made by the physician completing the initial

assessment, and that new clinical information may have affected CT ordering. The comfort rating does not reflect all of the complexities of real-world decision-making. However, the comfort rating was always completed by the same physician who estimated probability, and, because the case report form was completed at the time of initial assessment, the comfort rating was also unaffected by any later change in clinical status.

Because of the risks of radiation, it was not ethical to perform a CT for every subject.^{14,15} It is possible that some children with injuries that would have been visible on CT, but did not require intervention, were not identified. Practice patterns of individual surgeons may have affected whether subjects received intervention for certain injuries. Probability information was collected in categories rather than asking physicians to provide an estimate of probability. This practice probably increased the proportion of the probability estimation items that were completed but also limited our ability to evaluate finer estimates of probabilities. Our sample was restricted to Canadian pediatric emergency departments and may not be generalizable to settings outside of Canada or to emergency departments where both adults and children are seen. The data presented were collected between 2001 and 2005. Physician understanding about traumatic brain injury has changed substantially in the intervening years, which may impact the accuracy of predicted probabilities.

CONCLUSIONS

Physician estimates of probability of any brain injury in children were collectively accurate for children with low and moderate degrees of predicted risk. Risk was substantially underestimated in infants. Risk was overestimated in other circumstances, such as when physicians were asked to evaluate the probability of injury requiring intervention, or in the subset of children with very high risk of actual injury. Estimated risk varied between raters, indicating that some physicians were likely less accurate than others. Further research may identify

differences between effective and ineffective strategies for estimating risk, both in infants and older children. Both qualitative research that explores physicians' thought processes and quantitative research that describes real-life behaviour would be useful.

Acknowledgements: The authors thank the site investigators and other members of the Pediatric Emergency Research Canada (PERC) Head Injury Study Group for their contributions to study design and data collection: Keith Aronyk, MD (University of Alberta), Benoit Bailey (Hospital Ste-Justine), Laurel Chauvin-Kimoff (McGill University), Mark Hamilton, MD (University of Calgary), D. Anna Jarvis (University of Toronto), Gary Joubert (University of Western Ontario), Don McConnell (University of Alberta), Cheri Nijssen-Jordan (University of Calgary), Martin Pusic (Columbia University Medical Centre), Martin Reed, MD (University of Manitoba), Norm Silver (University of Manitoba), Ian Stiell (University of Ottawa), Brett Taylor (Dalhousie University), and Michael Vassilyadi, MD (University of Ottawa). We thank the following site study coordinators for their much appreciated assistance: Jennifer Spruyt, Eleanor Fitzpatrick, Rita Arsenault, Bev Irwin, Rose Jacobson, Sue Heathcote, Lanna Bryska, Nathalie Franc, Geri Siebenga St. Jean, and Diane Laforte. We also thank My-Linh Tran and Sheryl Domingo for data management, the physicians, nurses, and clerks at the study sites who voluntarily assisted with case identification and data collection, and the patients and families who participated.

Disclosure: Dr. Daymont's time was funded by the Children's Hospital Research Institute of Manitoba and the Manitoba Health Research Council. The CATCH Study was funded by peer-reviewed grants from the Canadian Institutes of Health Research (CIHR Funding Reference Number: MOP-43911), the Emergency Health Services Branch of the Ontario Ministry of Health, and the Alberta Children's Hospital Foundation.

Competing interests: None declared.

REFERENCES

- Poses RM, Cebul RD, Collins M, et al. The accuracy of experienced physicians' probability estimates for patients with sore throats. Implications for decision making. *JAMA* 1985 Aug 16;254(7):925-9.
- Poses RM, Anthony M. Availability, wishful thinking, and physicians' diagnostic judgments for patients with suspected bacteremia. *Med Decis Making* 1991 Sep;11(3):159-68.
- Poses RM, Bekes C, Copare FJ, et al. What difference do two days make? The inertia of physicians' sequential prognostic judgments for critically ill patients. *Med Decis Making* 1990;10(1):6-14.
- Schulman KA, Escarce JJ, Eisenberg JM, et al. Assessing physicians' estimates of the probability of coronary artery disease: the influence of patient characteristics. *Med Decis Making* 1992 Jun;12(2):109-14.
- Berlowitz DR, Ghalil K, Moskowitz MA. The use of follow-up chest roentgenograms among hospitalized patients. *Arch Intern Med* 1989 Apr;149(4):821-5.
- Smith WR, Poses RM, McClish DK, et al. Prognostic judgments and triage decisions for patients with acute congestive heart failure. *Chest* 2002 May;121(5):1610-7.
- Dolan JG, Bordley DR, Mushlin AI. An evaluation of clinicians' subjective prior probability estimates. *Med Decis Making* 1986;6(4):216-23.
- Croskerry P. Cognitive forcing strategies in clinical decision making. *Ann Emerg Med* 2003;41(1):110-20.
- Croskerry P. Clinical cognition and diagnostic error: applications of a dual process model of reasoning. *Adv Health Sci Educ Theory Pract* 2009;14(Suppl 1):27-35.
- Grimshaw JM, Eccles MP, Steen N, et al. Applying psychological theories to evidence-based clinical practice: identifying factors predictive of lumbar spine x-ray for low back pain in UK primary care practice. *Implement Sci* 2011; 6(1):55.
- Scott SD, Grimshaw J, Klassen TP, et al. Understanding implementation processes of clinical pathways and clinical practice guidelines in pediatric contexts: a study protocol. *Implement Sci* 2011 Dec 28;6(1):133.
- Osmond MH, Klassen TP, Wells GA, et al. CATCH: a clinical decision rule for the use of computed tomography in children with minor head injury. *CMAJ* 2010;182(4): 341-8.
- Kuppermann N, Holmes JF, Dayan PS, et al. Identification of children at very low risk of clinically-important brain injuries after head trauma: a prospective cohort study. *Lancet* 2009;374(9696):1160-70.
- Hall EJ, Brenner DJ. Cancer risks from diagnostic radiology. *Br J Radiol* 2008;81(965):362-78.
- Pearce MS, Salotti JA, Little MP, et al. Radiation exposure from CT scans in childhood and subsequent risk of leukaemia and brain tumours: a retrospective cohort study. *Lancet* 2012 Aug;380(9840):499-505.