Computed Tomography Findings for Intracerebral Hemorrhage Have Little Incremental Impact on Post-Stroke Mortality Prediction Model Performance

Darin B. Zahuranec a  Brisa N. Sánchez b  Devin L. Brown a  Jeffrey J. Wing b, c  Melinda A. Smith a  Nelda M. Garcia a  William J. Meurer a  Lewis B. Morgenstern a, c  Lynda D. Lisabeth a, c  

a Stroke Program, University of Michigan Health System, Departments of b Biostatistics and c Epidemiology, University of Michigan School of Public Health, Ann Arbor, Mich., USA

Key Words
Epidemiology • Cerebral infarction • Prediction of outcome • Intracerebral hemorrhage

Abstract

Background: Stroke outcome studies often combine cases of intracerebral hemorrhage (ICH) and ischemic stroke (IS). These studies of mixed stroke typically ignore computed tomography (CT) findings for ICH cases, though the impact of omitting these traditional predictors of ICH mortality is unknown. We investigated the incremental impact of ICH CT findings on mortality prediction model performance.

Methods: Cases of ICH and IS (2000–2003) were identified from the Brain Attack Surveillance in Corpus Christi (BASIC) project. Base models predicting 30-day mortality included demographics, stroke type, and clinical findings (National Institutes of Health Stroke Scale (NIHSS) +/– Glasgow Coma Scale (GCS)). The impact of adding CT data (volume, intraventricular hemorrhage, infratentorial location) was assessed with the area under the curve (AUC), unweighted sum of squared residuals (Ŝ), and integrated discrimination improvement (IDI). The model assessment was performed first for the mixed case of IS and ICH, and then repeated for ICH cases alone to determine whether any lack of improvement in model performance with CT data for mixed stroke type was due to IS cases naturally forming a larger proportion of the total sample than ICH.

Results: A total of 1,256 cases were included (86% IS, 14% ICH). Thirty-day mortality was 16% overall (11% for IS; 43% for ICH). When both clinical scales (NIHSS and GCS) were included, none of the model performance measures showed improvement with the addition of CT findings whether considering IS and ICH together (ΔAUC: 0.002, 95% CI –0.01, 0.02; ΔŜ: –3.0, 95% CI –9.1, 2.6; IDI: 0.017, 95% CI –0.004, 0.05) or considering ICH cases alone (ΔAUC: 0.02, 95% CI –0.02, 0.08; ΔŜ: –2.0, 95% CI –9.7, 3.4; IDI 0.065, 95% CI –0.03, 0.21). If NIHSS was the only clinical scale included, there was still no improvement in AUC or Ŝ when CT findings were added for the sample with IS/ICH combined (ΔAUC: 0.005, 95% CI –0.01, 0.02; ΔŜ: –5.0, 95% CI –11.6, 1.0) or for ICH cases alone (ΔAUC: 0.05, 95% CI –0.002, 0.11; ΔŜ: –4.2, 95% CI –11.5, 2.3). However, IDI was improved when NIHSS was the only clinical scale included—there was still no improvement in AUC or Ŝ when CT findings were added for the sample with IS/ICH combined (ΔAUC: 0.005, 95% CI –0.01, 0.02; ΔŜ: –5.0, 95% CI –11.6, 1.0) or for ICH cases alone (ΔAUC: 0.05, 95% CI –0.002, 0.11; ΔŜ: –4.2, 95% CI –11.5, 2.3). However, IDI was improved when NIHSS was the only clinical scale for IS/ICH combined (IDI: 0.029, 95% CI 0.002, 0.065) and ICH alone (IDI: 0.12, 95% CI 0.005, 0.26).

Conclusions: Excluding ICH CT findings had only minimal impact on mortality prediction model performance whether examining ICH and IS together or ICH alone. These findings have important implications for the design of clinical studies involving ICH patients.

Copyright © 2012 S. Karger AG, Basel

Karger AG, Basel
Fax +41 61 306 12 34
E-Mail karger@karger.ch
www.karger.com

© 2012 S. Karger AG, Basel
1015–9770/12/3408606/0$38.00/0
Accessible online at:
www.karger.com/ced

Darin B. Zahuranec, MD
University of Michigan Cardiovascular Center
1590 East Medical Center Drive, SPC #5855
Ann Arbor, MI 48109-5855 (USA)
Tel. +1 734 936 9075, E-Mail zdarin@umich.edu

© 2012 S. Karger AG, Basel
1015–9770/12/3408606/0$38.00/0
Accessible online at:
www.karger.com/ced
Introduction

Stroke outcome studies often combine cases of intracerebral hemorrhage (ICH) and ischemic stroke (IS) to increase sample size. However, accounting for mortality risk in these studies of mixed stroke type is potentially problematic. Computed tomography (CT) findings such as hemorrhage volume, location, and intraventricular blood have been widely reported as powerful predictors of mortality in ICH [1–3], though these findings have no direct correlate in cases of IS. Mortality risk estimates for an individual may therefore be inaccurate if CT data are ignored for ICH.

Many large-scale population-based or quality improvement studies, such as those based on the European Registers of Stroke [4] or American Heart Association Get with the Guidelines database [5] do not traditionally report data on ICH volume. These and other studies investigating mixed populations of IS and ICH have either reported results stratified by stroke type, used a single baseline severity scale such as the National Institutes of Health Stroke Scale (NIHSS) across stroke types, or simply adjusted for stroke type as hemorrhagic or ischemic in regression models [4–9]. It is unknown if methods of severity adjustment which ignore CT data in ICH are adequate. While CT scans are readily available in clinical settings, obtaining original CT images to code detailed findings such as ICH volume and location for research studies adds extra expense and effort, particularly for large multicenter studies.

The primary goal of this study was to determine the incremental value of adding CT findings for ICH cases to models predicting 30-day mortality in a population-based study combining ICH and IS. We compared the performances (discrimination and calibration) of models that contained demographics, exam findings, and stroke type to those that also incorporated CT data.

Methods

Case Identification

Data for this project came from the Brain Attack Surveillance in Corpus Christi (BASIC) project [10]. BASIC is a population-based stroke surveillance project in Nueces County, Tex., USA. Detailed methods have been previously described [10–12]. Briefly, active and passive surveillance are used to identify all cases of stroke in the county. Board certified neurologists validate all cases of stroke and determine stroke type. IS is defined as the acute onset of a focal neurological deficit attributable to a specific cerebrovascular distribution that persisted for greater than 24 h. ICH is defined as non-traumatic acute onset of focal neurological symptoms associated with a focal collection of blood within the brain parenchyma, excluding cases of hemorrhagic conversion of a cerebral infarction or tumor. The study population consisted of all BASIC patients with first-ever IS or ICH from January 1, 2000 through December 31, 2003. The time frame was restricted to these cases as ICH imaging data were available for this time period.

Data Collection and ICH Scan Review

Trained abstractors reviewed medical charts for key data. Initial NIHSS was abstracted from the chart by trained abstractors using a previously validated method [13] (97.5%) or obtained from the medical record (2.5%). Initial Glasgow Coma Scale (GCS) score was similarly obtained from the medical record (54.1%) or calculated based on information abstracted from the medical record (45.9%). Ethnicity was determined from the medical record and reported as either Mexican American or non-Hispanic white. Other ethnic groups were excluded from the analysis due to low numbers. CT scans of ICH patients were reviewed by study neurologists and coded for ICH volume, location as infratentorial or supratentorial, and presence or absence of intraventricular hemorrhage as previously described [14, 15]. These CT parameters were selected due to their known association with 30-day mortality [3]. ICH volume was calculated with the ABC/2 method [16]. Vital status at 30 days was determined from medical record review along with state and national databases as previously described [17]. Brain imaging studies were not available for detailed review for IS cases.

Statistical Methods

Baseline characteristics were compared by stroke type and by 30-day mortality status using t tests, χ² tests, or non-parametric rank sum tests. We developed multivariable logistic regression models predicting 30-day mortality with and without the following CT findings: ICH volume (continuous), intraventricular hemorrhage (yes/no), and infratentorial hemorrhage (yes/no). Because CT findings are not available for IS cases, they were included only with their interaction with the stroke type indicator variable denoting that the stroke case was ICH. The models were: (1) the base model which included demographic information (age, gender, ethnicity), NIHSS, stroke type (IS = 0 or ICH = 1), and the interaction between NIHSS and stroke type; (2) the base model with the addition of ICH CT parameters; (3) the base model with GCS (modeled as a quadratic term due to non-linear association with log-odds of mortality) and the interaction between GCS and stroke type as additional severity measures, and (4) model 3 with ICH CT parameters. We assessed models with NIHSS only or NIHSS and GCS, since GCS is not routinely assessed in IS and may be unavailable in some datasets.

Predictive accuracy of models with and without CT findings was compared with three different metrics. The area under the receiver operating characteristic curve (AUC) [18, 19] and the integrated discrimination improvement (IDI) [20] were used to assess model discrimination, while model calibration was assessed with the unweighted sum of squared residuals (S) [21]. Inference regarding improvement of models with CT findings was conducted by estimating the change in AUC and S (ΔAUC and ΔS), the IDI index, and their 95% CI. Models were considered to be significantly different from a comparison model if the 95% CIs did not include zero. Bootstrapping [22] was used to construct CIs and to correct for over-fitting (i.e. deriving the model and as-
Since ICH forms a naturally smaller subset of the total population, it is possible that lack of improvement in the model performance may be due to the fact that CT findings only improve the model fit for the minority of cases. To investigate this possibility, the assessment of model performance was repeated for ICH cases alone.

Statistical analysis was done in R version 2.9.2 and SAS version 9.2 (SAS Institute, Cary, N.C., USA). This project was approved by the Institutional Review Board of the University of Michigan and the individual Corpus Christi hospital systems.

**Results**

A total of 1,256 cases were included (IS 1,083 (86%), ICH 171 (14%)). Descriptive characteristics of the population overall, by 30-day mortality, and by stroke type are shown in Table 1. Most patients (IS 68%, ICH 74%) arrived to medical attention within 24 h of symptom onset. Compared to those with IS, patients with ICH were less likely to have diabetes, had a higher baseline NIHSS, and had a lower baseline GCS. Thirty-day mortality was 197/1,256 (16%) overall, 123/1,083 (11%) for IS, and 74/171 (43%) for ICH.

Parameter estimates for the multivariable logistic regression models predicting 30-day mortality are shown in online supplementary table E1 (for all online suppl. material, see www.karger.com/doi/10.1159/000339684). Table 2 demonstrates the fit parameters (AUC and Ŝ) and measures of improvement (ΔAUC, ΔŜ, and IDI) for these models. For the baseline model (model 1), which incorporated age, gender, ethnicity, NIHSS, stroke type, and the interaction between stroke type and NIHSS, the AUC was 0.83 (95% CI 0.78, 0.87) and Ŝ was 113.3 (95% CI 93.3, 135.0). Adding the ICH CT findings (model 2) resulted in no significant change in AUC (ΔAUC for model 2-model 1: 0.005, 95% CI –0.01, 0.02), and slight but non-significant improvement in Ŝ (ΔŜ for model 2-model 1: –5.0, 95% CI 11.6, 1.0). However, the IDI comparing model 2 to model 1 was significantly different from zero, 0.029 (95% CI 0.002, 0.065). The change in AUC and the IDI comparing model 2 and model 1 are shown in figure 1a, b. The area between the two receiver operating characteristic (ROC) curves (fig. 1a) is the ΔAUC, and the area between the sensitivity versus cutoff value curves (fig. 1b) is the IDI.

We also investigated a model that incorporated both NIHSS and GCS as measures of clinical severity (model 3). Adding CT data to model 3 did not result in any change in

### Table 1. Description of the population by 30-day mortality status and stroke type

<table>
<thead>
<tr>
<th>Characteristic</th>
<th>All patients (n = 1,256)</th>
<th>30-day mortality status</th>
<th>p</th>
<th>Stroke type</th>
<th>IS (n = 1,085)</th>
<th>ICH (n = 171)</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>deceased at 30 days (n = 197)</td>
<td>alive at 30 days (n = 1,059)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mean age (SD)</td>
<td>71.1 (12.5)</td>
<td>75.9 (13.1)</td>
<td>70.2 (12.2)</td>
<td>&lt;0.001</td>
<td>70.9 (12.5)</td>
<td>72.2 (12.9)</td>
<td>0.23</td>
</tr>
<tr>
<td>Female sex</td>
<td>655 (52)</td>
<td>122 (62)</td>
<td>533 (50)</td>
<td>0.003</td>
<td>566 (52)</td>
<td>89 (52)</td>
<td>0.98</td>
</tr>
<tr>
<td>Mexican American</td>
<td>670 (53)</td>
<td>81 (41)</td>
<td>589 (56)</td>
<td>&lt;0.001</td>
<td>572 (53)</td>
<td>98 (57)</td>
<td>0.26</td>
</tr>
<tr>
<td>Hypertension</td>
<td>860 (68)</td>
<td>129 (65)</td>
<td>731 (69)</td>
<td>0.33</td>
<td>741 (68)</td>
<td>119 (70)</td>
<td>0.73</td>
</tr>
<tr>
<td>Diabetes</td>
<td>452 (36)</td>
<td>60 (30)</td>
<td>292 (37)</td>
<td>0.08</td>
<td>403 (37)</td>
<td>49 (29)</td>
<td>0.03</td>
</tr>
<tr>
<td>Coronary artery disease</td>
<td>364 (29)</td>
<td>77 (39)</td>
<td>287 (27)</td>
<td>&lt;0.001</td>
<td>320 (30)</td>
<td>44 (26)</td>
<td>0.31</td>
</tr>
<tr>
<td>Hyperlipidemia</td>
<td>243 (19)</td>
<td>30 (15)</td>
<td>213 (20)</td>
<td>0.11</td>
<td>219 (20)</td>
<td>24 (14)</td>
<td>0.06</td>
</tr>
<tr>
<td>Atrial fibrillation</td>
<td>131 (10)</td>
<td>39 (20)</td>
<td>92 (9)</td>
<td>&lt;0.001</td>
<td>117 (11)</td>
<td>14 (8)</td>
<td>0.3</td>
</tr>
<tr>
<td>Median NIHSS (IQR)</td>
<td>4 (2, 8)</td>
<td>14 (6, 26)</td>
<td>3 (1, 6)</td>
<td>&lt;0.0001</td>
<td>3 (1, 7)</td>
<td>10 (3, 21)</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>Median GCS (IQR)</td>
<td>15 (14, 15)</td>
<td>10 (6, 14)</td>
<td>15 (14, 15)</td>
<td>&lt;0.0001</td>
<td>15 (14, 15)</td>
<td>12 (6, 15)</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>30-day mortality</td>
<td>197 (16)</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
</tr>
</tbody>
</table>

CT parameters for ICH patients only (n = 171)

<table>
<thead>
<tr>
<th>Characteristic</th>
<th>Median ICH volume in cc (IQR)</th>
<th>Intraventricular hemorrhage present</th>
<th>Infratentorial hemorrhage</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>11 (3, 32)</td>
<td>90 (53)</td>
<td>21 (12)</td>
</tr>
<tr>
<td></td>
<td>27 (9, 68)</td>
<td>56 (76)</td>
<td>14 (19)</td>
</tr>
<tr>
<td></td>
<td>6 (1, 15)</td>
<td>34 (35)</td>
<td>7 (7)</td>
</tr>
</tbody>
</table>

Figures are n (%) unless otherwise specified.
AUC \( AUC \) for model 4-model 3: 0.002, 95% CI \(-0.01, 0.02\) or \( \hat{S} \) (\( \hat{S} \) for model 4-model 3: \(-3.0, 95\% CI \(-9.1, 2.6\)), or improvement in the IDI (IDI for model 4-model 3: 0.017, 95% CI \(-0.004, 0.05\)). The change in AUC and the IDI comparing model 4 and model 3 are shown in figure 1c, d.

The analysis was then repeated restricted to ICH cases alone, with model parameter estimates shown in online suppl. table E2 and measures of improvement shown in table 3. Increasing ICH volume and infratentorial location of hemorrhage was associated with risk of death.
among ICH cases in both model 2 and model 4. Adding CT findings to the base model containing demographics and NIHSS (comparing model 2 to model 1) resulted in a slight but non-significant improvement in AUC and \( \hat{S} \), though IDI was improved. However, if GCS was added to the base model as an additional assessment of the clinical exam, there was no significant improvement in AUC, \( \hat{S} \), or IDI with the addition of CT findings (comparing model 4 to model 3). The changes in AUC and the IDI when restricting to ICH cases alone are shown in figure 2.

![Fig. 2. Model comparison curves for ICH cases only. The comparison of model 1 (demographics and NIHSS) to model 2 (adding CT findings) is shown in a: ROC curve demonstrating the change in AUC, and b sensitivity versus probability cutoff curve demonstrating the IDI. The comparison of model 3 (model 1 plus GCS) to model 4 (adding CT findings) is shown in c: ROC curve demonstrating the change in AUC, and d sensitivity versus probability cutoff curve demonstrating the IDI. Note that b and d demonstrate that sensitivity is actually decreased at small probability cutoffs when including CT findings, though the net IDI is positive and significant in b. When GCS is added to the model (d), the gain in sensitivity at larger probability cutoffs is attenuated, resulting in a non-significant IDI.](image)

### Table 3. Assessment of model performance: ICH cases only

<table>
<thead>
<tr>
<th>Model</th>
<th>Covariates</th>
<th>AUC (higher is better discrimination)</th>
<th>( \hat{S} ) (lower is better calibration)</th>
<th>IDI (&gt;0 is better discrimination)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>value</td>
<td>95% CI</td>
<td>value</td>
</tr>
<tr>
<td>1</td>
<td>Base model: age, gender, ethnicity, NIHSS</td>
<td>0.83</td>
<td>0.74, 0.90</td>
<td>30.9</td>
</tr>
<tr>
<td>2</td>
<td>Base model plus ICH CT findings*</td>
<td>0.87</td>
<td>0.80, 0.92</td>
<td>26.3</td>
</tr>
<tr>
<td></td>
<td>Change in AUC, ( \hat{S} ), or IDI (model 2-model 1)</td>
<td>0.05</td>
<td>-0.002, 0.11</td>
<td>-4.2</td>
</tr>
<tr>
<td>3</td>
<td>Base model plus GCS†</td>
<td>0.85</td>
<td>0.77, 0.91</td>
<td>28.8</td>
</tr>
<tr>
<td>4</td>
<td>Model 3 plus ICH CT findings*</td>
<td>0.87</td>
<td>0.80, 0.93</td>
<td>26.6</td>
</tr>
<tr>
<td></td>
<td>Change in AUC, ( \hat{S} ), or IDI (model 4-model 3)</td>
<td>0.02</td>
<td>-0.02, 0.08</td>
<td>-2.0</td>
</tr>
</tbody>
</table>

* CT findings for ICH were hemorrhage volume in cc, intraventricular hemorrhage (yes/no), and infratentorial hemorrhage (yes/no).

† Modeled as a quadratic term due to non-linear association with log-odds of mortality.
Discussion

We found that excluding CT data for ICH cases had little overall impact on model performance when predicting 30-day mortality in either a population of mixed ICH and IS, or when restricting to ICH cases alone. CT findings appeared to add relatively little benefit to model discrimination ability (AUC and IDI) or calibration (\( \hat{S} \)) beyond that provided by demographics, clinical stroke severity (NIHSS and GCS), and stroke type (when considering both ICH and IS). These findings have implications for stroke outcome studies where it may not be practical to have each CT scan individually coded, such as multicenter quality improvement projects, studies based on administrative data, or large-scale population-based studies with many events.

Our findings do not contradict the well-known association between mortality and CT findings such as hemorrhage volume or location in ICH [1]. Indeed, ICH volume and infratentorial location were associated with mortality in all models. Rather, these results suggest that depending on the particular research question at hand, it may not be necessary to go to the extra effort and expense of obtaining detailed CT data for ICH cases. For example, if the goal of a model is to determine which predictor variables have the strongest association with post-stroke mortality, then ICH CT findings should be included. However, if the goal is to account for stroke mortality risk while investigating the impact of other factors (such as gender or adherence to quality indicators) on stroke outcome, CT data may not be necessary. Cost of obtaining the CT data may vary substantially depending on the study setting (e.g. single center vs. multicenter or population-based) and our results may help investigators to assess the value of obtaining CT data in their own studies.

The lack of influence of CT findings even when restricting to ICH cases alone may seem surprising when considering that the majority of existing ICH predictive models include CT findings as predictors [1]. However, this finding is not unprecedented: the Essen ICH score is a validated ICH predictive model based on age, level of consciousness, and NIHSS that does not include imaging findings and yet has been shown to compare favorably to other models which do incorporate CT findings [23]. It appears that substantial information about mortality risk is contained in demographic variables and clinical examination findings, and therefore addition of CT findings adds relatively little to model performance. We cannot completely exclude the possibility of a type II error, particularly with the reduction in sample size when restricting to ICH cases alone. However, the point estimates of the change in AUC and \( \hat{S} \) in table 3 suggest that the absolute incremental benefit of CT data on model performance is relatively small.

Our results differed slightly depending on whether NIHSS alone or NIHSS plus GCS were included as measures of clinical severity. When considering the model with NIHSS as the only clinical severity measure (model 1), we found that there was an improvement in IDI, but not AUC or \( \hat{S} \) with addition of CT findings. While AUC and IDI are both measures of the model’s ability to discriminate patients likely to die from those likely to survive, disparate results can occur due to the underlying characteristics of these measures. Both AUC and IDI can be viewed as the model sensitivity averaged across the range of probability thresholds (i.e. a probability value above which one might ‘classify’ a patient as more likely to die). However, AUC is weighted to give relatively more importance to sensitivity values at lower probability thresholds, whereas the IDI gives equal weight to all probability thresholds [20]. Because AUC gives more weight at low threshold values, the effect of the seemingly small loss of sensitivity at low threshold values (shown by the fact that the sensitivity curve with CT findings in figure 2b is below the curve for the model without CT findings at low cutoff values) is given more importance when calculating AUC compared to IDI. Therefore, the disparate results for AUC and IDI when comparing model 2 to model 1 may be reflective of CT findings improving model discrimination at higher probability thresholds.

This study has several limitations. Our results may not apply to other populations, or when considering outcomes other than 30-day mortality. We do not have data available on long-term functional outcome, such as 90-day modified Rankin Scale, for this cohort. While BASIC has a higher proportion of Mexican Americans than many other stroke populations, the 30-day mortality and relative proportion of IS and ICH are similar to prior reports from other populations [24]. We focused on 30-day mortality given the powerful and consistent association with ICH CT findings that has been extensively reported in the literature [1]. We did not review brain imaging studies for IS cases; however, imaging findings are not felt to be as useful a predictor of outcome for IS as for ICH [25]. We also did not include data on comorbid illness such as coronary artery disease. Adding additional predictors may have altered the values of the AUC, \( \hat{S} \), and IDI, but would be unlikely to change conclusions about the change in these parameters with or without CT findings.
Acknowledgements

This study was funded by the NIH and the National Institute of Neurological Disorders and Stroke (R01 NS38916). Dr. Zahuranec is supported by NIH grant K23 AG038731 from the National Institute on Aging.

References


26. The authors have no conflicts of interest to disclose.