Intercenter variance in clinical trials of head trauma—experience of the National Acute Brain Injury Study: Hypothermia

Guy L. Clifton, M.D., Sung C. Choi, Ph.D., Emmy R. Miller, R.N., Ph.D., Harvey S. Levin, Ph.D., Kenneth R. Smith, Jr., M.D., J. Paul Muizelaar, M.D., Ph.D., Franklin C. Wagner, Jr., M.D., Donald W. Marion, M.D., and Thomas G. Luerssen, M.D.

Department of Neurosurgery, Vivian L. Smith Center for Neurologic Research, University of Texas–Houston Health Science Center, Houston, Texas; Departments of Biostatistics and Neurosurgery, Medical College of Virginia/Virginia Commonwealth University, Richmond, Virginia; Department of Physical Medicine and Rehabilitation, Baylor College of Medicine, Houston, Texas; Department of Neurosurgery, St. Louis University, St. Louis, Missouri; Department of Neurological Surgery, University of California at Davis, Sacramento, California; Department of Neurosurgery, The Brain Trauma Research Center, University of Pittsburgh, Pennsylvania; and Department of Neurosurgery, Indiana University, Indianapolis, Indiana

Object. In a recently conducted trial of hypothermia in patients with severe brain injury, differences were found in the effects of hypothermia treatment among various centers. This analysis explores the reasons for such differences.

Methods. The authors reviewed data obtained in 392 patients treated for severe brain injury. Prerandomization variables, critical physiological variables, treatment variables, and accrual methodologies were investigated among various centers. Hypothermia was found to be detrimental in patients older than the age of 45 years, beneficial in patients younger than 45 years of age in whom hypothermia was present on admission, and without effect in those in whom normothermia was documented on admission. Marginally significant differences (p<0.054) in the intercenter outcomes of hypothermia-treated patients were likely the result of wide differences in the percentage of patients older than 45 years of age and in the percentage of patients in whom hypothermia was present on admission among centers. The trial sensitivity was likely diminished by significant differences in the incidence of mean arterial blood pressure (MABP) less than 70 mm Hg (p<0.001) and cerebral perfusion pressure (CPP) less than 50 mm Hg (p<0.05) but not intracranial pressure (ICP) greater than 25 mm Hg (not significant) among patients in the various centers. Hours of vasopressor usage (p<0.03) and morphine dose (p<0.001) and the percentage of dehydrated patients varied significantly among centers (p<0.001). The participation of small centers increased intercenter variance and diminished the quality of data.

Conclusions. For Phase III clinical trials we recommend: 1) a detailed protocol specifying fluid and MABP, ICP, and CPP management; 2) continuous monitoring of protocol compliance; 3) a run-in period for new centers to test accrual and protocol adherence; and 4) inclusion of only centers in which patients are regularly randomized.

Key Words • head injury • hypothermia • clinical trial • center variance

Abbreviations used in this paper: CPP = cerebral perfusion pressure; GCS = Glasgow Coma Scale; ICP = intracranial pressure; MABP = mean arterial blood pressure; SD = standard deviation; TISS = Therapeutic Intervention Scoring System.
in a National Institute of Neurological Disorders/National Institutes of Health–supported multicenter trial of hypothermia (33°C for 48 hours) for treatment of severe head injury. The trial began in July 1994 with the goal of enrolling 500 patients. Accrual ceased as a result of a futility analysis in May 1998.43 The analysis was not planned but suggested by the Performance and Safety Monitoring Board at the second interim analysis. The primary outcome measure was the score determined using the dichotomous Glasgow Outcome Scale at 6 months postinjury.10 The inclusion criteria were an admission GCS score of 3 to 8, age 16 to 65 years, and a nonpenetrating head injury. Patients were stratified by center and admission motor scores (1 and 2 compared with 3 and higher). Following previously published guidelines,1 our management protocol targeted CPP greater than 70 mm Hg. Dehydration was to be avoided. The final enrollment was 392 patients in two groups: standard management with hypothermia or standard management with normothermia.

In this trial hypothermia was found to reduce the percentage of patients with ICP above 30 mm Hg (42% in hypothermia- and 61% in normothermia-treated patients [p < 0.001]). In 52 patients older than 45 years of age, hypothermia was associated with increased poor outcomes (89% in hypothermia- and 69% in normothermia-treated patients [p < 0.08]). In 81 patients younger than age 45 years in whom hypothermia was documented on admission (≤ 35°C), hypothermia was associated with improved outcome (52% in hypothermia- and 76% in normothermia-treated patients [p < 0.02]). In 264 patients who were normothermic on admission there was no treatment effect.3

Eleven centers participated in the trial, although not all centers were involved for the entire study period. Deficiencies in patient accrual, protocol violations, and poor data quality forced the termination of four low-enrollment centers within the 1st year of the study. Two low-enrollment centers joined the study in its last year and had low overall accrual as a result of their late entry. The number of patients accrued by each center ranged from four to 96. Five high-enrollment centers accrued 88% of all randomized patients. In these five centers we examined the data for three sources of variance: 1) prerandomization, 2) physiological, and 3) treatment variables.

**Statistical Methods**

In the five largest centers, all prerandomization, treatment, and physiological variables found to have a significant effect on the primary outcome measure were examined. A variable with no effect on outcome would not contribute to intercenter variance in treatment effect. The intercenter differences in age and GCS score were tested using the analysis of variance and the rank-sum test, respectively. The intercenter differences in percentage of patients treated with hypothermia were analyzed by chi-square test. The incidence of MABP less than 70 mm Hg, CPP less than 50 mm Hg, and ICP greater than 25 mm Hg in each treatment group was examined using the multiple logistic regression test for possible multivariate effect, adjusting for age and GCS score. Similarly, data on intercenter variance with regard to other variables were analyzed using multiple logistic regression and the analysis of covariance, adjusting for age and GCS score when appropriate.23 The likelihood ratio test was used to determine the significance probability (p values) of overall center difference. Some continuous variables were summarized as the mean ± SD and tested using a simple t-test. The Mantel–Haenszel test was also used.11

**Results**

The differences in patient outcomes at 6 months postinjury among the centers are shown in Fig. 1. For the six low-enrollment centers, poor outcome in both hypothermia- and normothermia-treated groups varied widely. For the five high-enrollment centers, the intercenter poor-outcome variability was less, ranging from 51 to 70% in the hypothermia- and 33 to 74% in the normothermia-treated groups. The differences in percentage of poor outcomes among the five largest centers were not significant in the latter group but were marginally significant in the hypothermia-treated group (p < 0.054). At Centers 1 and 3 an apparent positive treatment effect was found, whereas at Centers 2, 4, and 5 an adverse effect of hypothermia was reported. A beneficial effect of hypothermia was shown at all five high-enrollment centers in patients 45 years of age or younger in whom hypothermia was present on admission, whereas a detrimental effect of the same treatment was observed in patients older than 45 years of age.

**TABLE 1**

<table>
<thead>
<tr>
<th>Center</th>
<th>No. of Patients</th>
<th>Mean Age (yrs) ± SD</th>
<th>Percentage &gt;45 Years of Age</th>
<th>Percentage Hypothermic on Admission</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>96</td>
<td>31 ± 12</td>
<td>9</td>
<td>24</td>
</tr>
<tr>
<td>2</td>
<td>73</td>
<td>32 ± 13</td>
<td>16</td>
<td>27</td>
</tr>
<tr>
<td>3</td>
<td>69</td>
<td>32 ± 14</td>
<td>19</td>
<td>36</td>
</tr>
<tr>
<td>4</td>
<td>77</td>
<td>30 ± 12</td>
<td>12</td>
<td>22</td>
</tr>
<tr>
<td>5</td>
<td>30</td>
<td>33 ± 11</td>
<td>17</td>
<td>17</td>
</tr>
</tbody>
</table>

* Intercenter differences in mean age, percentage of patients older than 45 years of age, and percentage who were hypothermic on admission were not significant.
Intercenter variance in clinical trials of head injury

Baseline Variables

Baseline variables were those observed before patients were randomized in a trial. The most important of these reported to affect outcome are GCS score and age. Age and the presence of hypothermia on admission most significantly influenced the effects of treatment with hypothermia. In particular, in the subgroup of the patients defined by age younger than 45 years and hypothermia on admission (≤ 35°C) markedly reduced percentages of poor outcomes were shown in the hypothermia treatment group (p < 0.02), adjusting for age and the GCS score. The mean and SD of age, percentage of patients older than 45 years of age, and the percentage of patients who were hypothermic on admission are shown by center in Table 1. The differences in distribution of these variables were not significant but the range was wide.

Physiological Variables

The physiological variables were recorded hourly for the first 96 hours after patients were randomized. The variables for which a single episode was found to exert a significant effect on outcome were the occurrence of MABP less than 70 mm Hg (p < 0.01), CPP less than 50 mm Hg (p < 0.01), and ICP greater than 25 mm Hg (p < 0.04). These variables had the same effects on outcome in patients in both treatment groups. The intercenter differences in the distributions of patients’ MABP, CPP, and ICP values are presented in Table 2. There were significant intercenter differences in the distribution of low MABP and CPP but not in elevated ICP. Center 3 had the smallest and Centers 2 and 5 had the largest proportion of patients with MABP less than 70 mm Hg and CPP less than 50 mm Hg. In Centers 2 and 3, a higher incidence of MABP less than 70 mm Hg was shown in normothermia-treated patients than in hypothermia-treated patients, whereas in Centers 1, 4, and 5 as well as in the overall group, a higher incidence of hypotension was observed in hypothermia-treated patients. These variations in low MABP and CPP, however, did not correlate with intercenter differences in hypothermia treatment effect.

Treatment Variables

Treatments other than ventricular drainage used to manage ICP included the provision of mannitol, hyperventilation, morphine, and muscle relaxant agents. Treatments used to manage MABP and CPP were fluid balance and vasopressor agents in addition to modalities for ICP control. The mean mannitol dose, cumulative 96-hour fluid balance, occurrence of PaCO₂ less than 30 mm Hg, and dose of vecuronium did not vary significantly among centers. There was significant variation in the percentage of patients who were hypothermic among centers in the mean time to reach the target temperature, however, was correlated to the outcome (p < 0.05). The intensity of management did not vary among centers. The TISS consists of 49 items with assigned values that reflect the intensity of intervention. Higher scores indicate more interventions. The mean TISS value for the first 4 days in the intensive care unit (calculated without the indicator for hypothermia) was 48 ± 7 for hypothermia- and 47 ± 8 for normothermia-treated patients (p < 0.02; Table 3). The greatest intercenter variability was demonstrated in the hypothermia-treated patients in whom vasopressor use ranged from 60 to 92%, whereas, in normothermia-treated patients it ranged from 65 to 74%. Patients received dopamine (55%), phenylephrine (43%), dobutamine (18%), and norepinephrine bitartrate (levophed; 16%). The morphine dose administered in each center was almost identical for both treatment groups. In hypothermia-treated patients, centers at which high morphine doses were administered also used more vasopressors (p < 0.06).

Time From Injury to Cooling and Time to Target Temperature

In the hypothermia treatment group, cooling began immediately after the patients were randomized. The range among centers in the mean estimated time from injury to cooling was 3.9 ± 2.2 to 4.6 ± 0.9 hours; the intercenter differences were marginally significant (p < 0.06). The range among centers in the mean time to reach 33°C post-injury was 7.5 ± 2.2 to 9.6 ± 4.1 hours; the differences were also marginally significant (p < 0.06). Neither the time to initiate cooling nor the time to reach the target temperature, however, was correlated to the outcome (p > 0.3).

---

*Hypo = percentage of patients in the hypothermia treatment group; normo = percentage of patients in the normothermia treatment group.
† p < 0.001 among centers.
‡ p < 0.05 among centers.
§ No significant intercenter difference.

**Table 2: Incidence of critical MABP, CPP, and ICP stratified by center**

<table>
<thead>
<tr>
<th>Center No.</th>
<th>No. of Patients</th>
<th>MABP &lt;70 mm Hg†</th>
<th>CPP &lt;50 mm Hg‡</th>
<th>ICP &gt;25 mm Hg§</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Hypo</td>
<td>Normo</td>
<td>Hypo</td>
<td>Normo</td>
</tr>
<tr>
<td>1</td>
<td>96</td>
<td>58</td>
<td>46</td>
<td>52</td>
</tr>
<tr>
<td>2</td>
<td>73</td>
<td>62</td>
<td>72</td>
<td>49</td>
</tr>
<tr>
<td>3</td>
<td>69</td>
<td>29</td>
<td>38</td>
<td>37</td>
</tr>
<tr>
<td>4</td>
<td>77</td>
<td>47</td>
<td>46</td>
<td>47</td>
</tr>
<tr>
<td>5</td>
<td>30</td>
<td>93</td>
<td>67</td>
<td>53</td>
</tr>
</tbody>
</table>

**Table 3: Summary of vasopressor use and mean morphine dose by center**

<table>
<thead>
<tr>
<th>Center No.</th>
<th>No. of Patients</th>
<th>Vasopressor Usage*</th>
<th>Morphine Dose (mg)†</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Hypo</td>
<td>Normo</td>
<td>Hypo</td>
</tr>
<tr>
<td>1</td>
<td>96</td>
<td>90</td>
<td>69</td>
</tr>
<tr>
<td>2</td>
<td>73</td>
<td>92</td>
<td>72</td>
</tr>
<tr>
<td>3</td>
<td>69</td>
<td>60</td>
<td>65</td>
</tr>
<tr>
<td>4</td>
<td>77</td>
<td>79</td>
<td>74</td>
</tr>
<tr>
<td>5</td>
<td>30</td>
<td>87</td>
<td>67</td>
</tr>
</tbody>
</table>

* p < 0.03 among centers.
† p < 0.001 among centers. Values are expressed as the means ± SD.
TABLE 4

<table>
<thead>
<tr>
<th>Center</th>
<th>No. of Patients Enrolled</th>
<th>No. W/ Poor Outcome</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Treatment</td>
<td>Control</td>
</tr>
<tr>
<td>A</td>
<td>60</td>
<td>60</td>
</tr>
<tr>
<td>B</td>
<td>50</td>
<td>50</td>
</tr>
<tr>
<td>C</td>
<td>45</td>
<td>45</td>
</tr>
<tr>
<td>D</td>
<td>40</td>
<td>40</td>
</tr>
<tr>
<td>E</td>
<td>5</td>
<td>5</td>
</tr>
</tbody>
</table>

* Although the difference between treatment and control groups is significant (p = 0.043) when analyzed without data from Center E, no significance (p = 0.073) is shown once Center E data are included in the analysis.

Large High-Enrollment Compared With Small Low-Enrollment Centers

In the trial, five large centers enrolled 88% of the 392 randomized patients, and the remaining 12% were accrued by six small centers. The differences between the five large centers and the six small centers were compared. First, the proportion of missing data and the quality of the data in general were related to the center size. For example, the admission GCS score was missing for 36% of the 47 patients accrued by the small centers, whereas it was recorded for every patient accrued by the five large centers. The MABP and fluid balance data were missing for 11% and 13% of patients, respectively, in the small centers, whereas the corresponding data were complete for all patients in the large centers.

The number of patients randomized by five of the six small centers ranged from four to nine, and information obtained in only 30 of 47 patients enrolled at the small centers could be analyzed because of missing data. Thus, the comparison between the two types of (small- and large-sized) centers lacked statistical power because of the small sample size. Even so, the difference in percentage of poor outcome between the two types of centers was significant (59% at large and 37% at small centers in hypothermia-treated patients; 57% at large and 55% at small centers in normothermia-treated patients [p < 0.04]). The differences in the mean morphine dose were also significant: 8.3 ± 4.4 mg at large and 7.2 ± 3 mg at small centers in hypothermia-treated patients; 8.4 ± 4.8 mg at large and 5.9 ± 2 mg at small centers in normothermia-treated patients (p < 0.05).

Discussion

It may be speculated that the intercenter difference in treatment effect was probably, in part, a result of the distribution of patients older than 45 years of age and patients in whom hypothermia was present on admission. An increased percentage of older patients would have increased the probability of a detrimental effect of hypothermia. On the other hand, a greater number of hypothermic patients on admission would have increased the probability of finding a beneficial effect. Treatment with hypothermia only at Centers 1 and 3 had an overall treatment effect (GCS Scores 3–8). Center 1 had the lowest percentage (10%) of patients older than 45 years of age and Center 3 had the highest percentage (36%) of patients in whom hypothermia was present on admission. Center 5, at which the most adverse effects of hypothermia were observed, had the lowest percentage (17%) of hypothermic patients on admission and a high percentage (17%) of patients older than 45 years of age. These findings suggest that differences in these baseline variables were a major factor in intercenter differences in treatment effect.

Despite the range in incidence of MABP less than 70 mm Hg and CPP less than 50 mm Hg among centers, intercenter and intergroup variations within centers did not correlate with center treatment effect. Even so, it would be naïve to conclude that the low MABP and CPP have no effect on intercenter differences, because of complex multifactorial interaction effects of these with other variables. The significant variations in the incidence of critically low MABP and CPP among centers is no surprise given the variations in the use of morphine, vasopressor agents, and fluids. Vasopressor agents were used with varying hemodynamic effects: dopamine primarily increases cardiac output; dobutamine increases cardiac output and is a vasodilator; phenylephrine decreases cardiac output with mild vasoconstriction; and norepinephrine bitartrate is a potent vasoconstrictor.14,15,17,18 Morphine is a dose-dependent vasodilator, and significant variation in mean dose was demonstrated among centers.16 Dehydration produces vasoconstriction in some physiological states, and the incidence of dehydration differed among centers.8,9 The lack of a standard method for the management of fluids, morphine, and vasopressor therapy probably led to the variations in MABP and CPP management.

Of the three sources of center variance, statistical adjustments for the differences in the baseline variables such as age and GCS score are justified because the variables were measured before the treatment randomization. This type of adjustment increases precision in clinical trials by removing the effects of a known source of variance.2,6 Adjustments for differences in physiological or treatment variables are likely to yield misleading results. In this study, for example, in 71% of patients in the normothermia treatment group at least one hourly ICP measurement that exceeded 25 mm Hg was demonstrated compared with 59% of patients in the hypothermia treatment group. If outcomes were improved by hypothermia, adjustment for ICP differences would incorrectly remove the positive effect.

Six of the 11 centers enrolled fewer than 14 patients. Low enrollment at centers is likely to cause imbalance between the two arms of a trial. The possible effect of small centers on statistical power can be illustrated using a hypothetical example. Consider a five-center trial based on the total sample size of 400 severely head injured patients (Table 4). Centers A, B, C, and D accrued 80 to 120 patients and randomized them into two study arms. Center E accrued only five patients in each of the two groups, with a difference of only two patients with poor outcomes between the two groups. For the high-enrollment centers, which had accrued over 97% of the patients in the trial, poor outcome in the treatment group is significantly less than in the control group (p < 0.05). When the outcomes from the low-enrollment centers are included in the analysis, however, the overall difference between the two groups is no longer significant (p > 0.07). In a low-enrollment center, the possibility of observing results quite different
Intercenter variance in clinical trials of head injury

from those achieved at large centers is high, as illustrated by results of the National Acute Brain Injury Study: Hypothermia (Fig. 1). Furthermore, it has been reported that, in general, the quality of data in small centers is lower than in centers with high patient accrual.\textsuperscript{7,13} The same pattern of low data quality was found in centers with low accrual in this trial.

Conclusions

The trial was conducted using detailed management protocols. Therefore, we were concerned that significant intercenter differences were found with respect to outcomes and other variables that are difficult to explain by chance alone. The intercenter variance in baseline as well as physiological and treatment variables can increase the variation in outcomes and reduce the overall sensitivity of the trial. In conclusion, based on the results of this trial, we recommend: 1) a very detailed management protocol specifying MABP, CPP, and fluid management; 2) continuous monitoring of compliance with the management protocol; 3) a run-in period for new centers before beginning accrual in a trial; and 4) participation only by centers at which patients are regularly randomized. The inclusion of small centers, at which the extent of missing data is considerable and which vary significantly from large centers, could actually harm a trial rather than help it.

References

   344:556–563, 2001
   1031–1034, 1976
   22:719–748, 1959
   305:852–855, 1981
   St Louis: Mosby–Year Book, 1998, Section II, p 739

Manuscript received November 3, 2000. Accepted in final form June 27, 2001. The work in this report was supported by funding from the National Institutes of Health (Subcontract No. 5 RO1 NS32786-06) to G. L. Clifton and S. C. Choi.

Address reprint requests to: Guy L. Clifton, M.D., Department of Neurosurgery, University of Texas Health Science Center, 6331 Fannin Street, Suite 7.148, Houston, Texas 77030, email: guy.l.clifton@uth.tmc.edu.