Clinical Trials in Head Injury

RAJ K. NARAYAN[1],[†]
1*Department of Neurosurgery, Temple University Hospital, Philadelphia, Pennsylvania*

MARY ELLEN MICHEL[2],[†]
2*National Institute of Neurological Disorders and Stroke, Bethesda, Maryland*

Beth Ansell[3],[‡], **Alex Baethmann**[4],[‡], **Anat Biegon**[5],[*],[‡], **Michael B. Bracken**[6],[*],[‡], **M. Ross Bullock**[7],[*],[‡], **Sung C. Choi**[7],[*],[‡], **Guy L. Clifton**[8],[*],[‡], **Charles F. Contant**[9],[*],[‡], **William M. Coplin**[10],[‡], **W. Dalton Dietrich**[11],[‡], **Jamshid Ghajar**[12],[*],[‡], **Sean M. Grady**[13],[‡], **Robert G. Grossman**[9],[‡], **Edward D. Hall**[14],[*],[‡], **William Heetderks**[15],[‡], **David A. Hovda**[16],[‡], **Jack Jallo**[17],[‡], **Russell L. Katz**[18],[*],[‡], **Nachshon Knoller**[19],[*],[‡], **Patrick M. Kochanek**[20],[‡], **Andrew I. Maas**[21],[*],[‡], **Jeannine Majde**[22],[‡], **Donald W. Marion**[20],[‡], **Anthony Marmarou**[7],[*],[‡], **Lawrence F. Marshall**[23],[*],[‡], **Tracy K. McIntosh**[13],[*],[‡], **Emmy Miller**[8],[‡], **Noel Mohberg**[24],[‡], **J. Paul Muizelaar**[25],[*],[‡], **Lawrence H. Pitts**[26],[‡], **Peter Quinn**[12],[‡], **Gad Riesenfeld**[27],[‡], **Claudia S. Robertson**[9],[*],[‡], **Kenneth I. Strauss**[28],[‡], **Graham Teasdale**[29],[*],[‡], **Nancy Temkin**[30],[*],[‡], **Ronald Tuma**[28],[‡], **Charles Wade**[31],[*],[‡], **Michael D. Walker**[15],[‡], **Michael Weinrich**[3],[‡], **John Whyte**[17],[*],[‡], **Jack Wilberger**[32],[‡], **A. Byron Young**[33],[*],[‡], and **Lorraine Yurkewicz**[34],[*],[‡]
3*NIH/NCMRR, Bethesda, Maryland*

4*Ludwig-Maximilians University of Munich, Munich, Germany*

5*Lawrence Berkeley National Laboratory, Berkeley, California*

6*Yale University School of Medicine, New Haven, Connecticut*

7*Virginia Commonwealth University, Richmond, Virginia*

8*University of Texas HSC, Houston, Texas*

9*Baylor College of Medicine, Houston, Texas*

10*Wayne State University School of Medicine, Detroit, Michigan*

11*University of Miami School of Medicine, Miami, Florida*

12*Brain Trauma Foundation, New York, New York*

13*Hospital of the University of Pennsylvania, Philadelphia, Pennsylvania*

14*Pfizer Global R&D, Ann Arbor, Michigan*

15*NIH/NINDS, Bethesda, Maryland*

16*University of California, Los Angeles, California*

17*Temple University School of Medicine, Philadelphia, Pennsylvania*

18*Food and Drug Administration, Rockville, Maryland*

19*Sheba Medical Center, Tel-Hashomer, Israel*

20*University of Pittsburgh School of Medicine, Pittsburgh, Pennsylvania*

21*Academic Hospital, Rotterdam, The Netherlands*

22*Office of Naval Research, Arlington, Virginia*

23*University of California, San Diego, La Jolla, California*

24*Parke-Davis Pharmaceutical Research, Ann Arbor, Michigan*

25*University of California at Davis Medical Center, Sacramento, California*

26*University of California, San Francisco, California*

27*Pharmos Corporation, Iselin, New Jersey*

28*Temple University School of Medicine, Philadelphia, Pennsylvania*

29*Southern General Hospital, Glasgow, Scotland*

30*University of Washington, Seattle, Washington*

31*Sausalito, California*

32*Allegheny General Hospital, Pittsburgh, Pennsylvania*

33*University of Kentucky Chandler Medical Center, Lexington, Kentucky*

34*Pfizer Pharmaceuticals, Groton, Connecticut.*

Abstract

Traumatic brain injury (TBI) remains a major public health problem globally. In the United States the incidence of closed head injuries admitted to hospitals is conservatively estimated to be 200 per 100,000 population, and the incidence of penetrating head injury is estimated to be 12 per 100,000, the highest of any developed country in the world. This yields an approximate number of 500,000 new cases each year, a sizeable proportion of which demonstrate signficant long-term disabilities. Unfortunately, there is a paucity of proven therapies for this disease. For a variety of reasons, clinical trials for this condition have been difficult to design and perform. Despite promising pre-clinical data, most of the trials that have been performed in recent years have failed to demonstrate any significant improvement in outcomes. The reasons for these failures have not always been apparent and any insights gained were not always shared. It was therefore feared that we were running the risk of repeating our mistakes. Recognizing the importance of TBI, the National Institute of Neurological Disorders and Stroke (NINDS) sponsored a workshop that brought together experts from clinical, research, and pharmaceutical backgrounds. This workshop proved to be very informative and yielded many insights into previous and future TBI trials. This paper is an attempt to summarize the key points made at the workshop. It is hoped that these lessons will enhance the planning and design of future efforts in this important field of research.

Keywords

clinical trials; head injury; intracranial pressure; outcome measures; traumatic brain injury; trial design; uniformed consent

INTRODUCTION

Tʀᴀᴜᴍᴀᴛɪᴄ ʙʀᴀɪɴ ɪɴᴊᴜʀʏ (TBI) remains a major public health problem. The incidence of all closed head injuries admitted to hospitals is conservatively estimated to be 200 per 100,000 population in the United States. This yields an approximate number of 500,000 cases in the United States

alone. Of these, 10% are generally classified at admission as severe (Glasgow Coma Scale [GCS] # 8), another 10% as moderate (GCS 9-12), and the rest as mild (GCS 13-15). Of the 50,000 patients who suffer a severe TBI, approximately one-third die even in the best of centers. Thus, in the United States, it is estimated that at least 17,500 patients die annually as a result of TBI, even without accounting for the fewer deaths among the moderate and the mild TBI groups. Of the survivors, a sizeable fraction demonstrates significant long-term disability.

The incidence of penetrating head injury in the United States is estimated to be 12 per 100,000, the highest of any developed country in the world. Approximately 40% of battlefield fatalities in the Vietnam War were due to head and neck wounds. While only 3% of the casualties admitted to hospitals during this conflict did not survive, 40% of these deaths were due to head injuries.

Despite the obvious public health implications of these numbers, there is a paucity of proven therapies for this disease. Furthermore, due to the complexity of the factors that impact upon the outcome from TBI, clinical trials for this condition have been difficult to design and conduct. Several trials that have been performed in recent years have failed to demonstrate significant improvement in outcomes, despite promising preclinical data. The reasons for failure were not always apparent and any insights gained were not always shared. It was feared that we in the head injury field were running the risk of repeating our mistakes.

Recognizing the importance of TBI, the National Institute of Neurological Disorders and Stroke (NINDS) sponsored a workshop in May 2000 that brought together experts from clinical, research, and pharmaceutical backgrounds. The goals of the workshop were to review previous clinical trials and to glean lessons that could be applied to future studies. The workshop was very successful in facilitating a free and lively exchange of ideas across disciplines and across commercial boundaries. This paper is broadly based on the proceedings of that meeting. However, the material has been edited and reorganized to make it more cohesive and to eliminate redundancies and ambiguities. Although not all participants presented papers, each one was actively involved in the discussions. All participants have recently reviewed the manuscript, added key references, and included any recent developments in the field, in order to make the information current and valuable. Any errors or omissions represent the failings of the authors who have attempted to faithfully yet concisely summarize a large volume of information and opinions. While the reader will obviously make the final judgment, we believe that the manuscript contains very valuable insights that are lucidly presented. It is hoped that this exercise has allowed us to learn from past experiences so that we may design the best possible studies in the future. We owe this to our patients, our field, and ourselves.

EXECUTIVE SUMMARY

Several issues related to the design and conduct of clinical trials in TBI were discussed in the workshop. While it is impractical to recapitulate all of the issues or all of the lessons learned, some of the key recommendations are summarized below. We must realize that the failure of TBI trials to date may not be due to poor trial design alone, but perhaps due to ineffective therapies or the selection of inappropriate target mechanisms. We must distinguish between these possibilities.

Identify and target specific mechanisms of cellular injury

- Obtain better understanding of the pathophysiology of TBI through continued research

- Target study drug or therapy to a mechanism that is known to occur in TBI

- Establish that a drug/therapy has the desired effect on the mechanism of injury *in vivo*

Obtain adequate preclinical data

- Model in the lab the severity of injury to be tested clinically

- Establish animal intensive care units (ICUs) to study severe TBI

- Study pharmacokinetics in multiple animal models to establish the dose, how soon, how frequently, and for how long to give drug or initiate and continue therapy

- Test the intervention in at least two rodent models (maybe two models of TBI such as weight drop or fluid percussion, or similar models in two rodent species) and in a larger animal if possible

- Test the intervention in more than one lab

- Try to model for both diffuse and focal injury, subarachnoid hemorrhage and ischemia

- Ensure adequate transport of a pharmaceutical agent into the cerebrospinal fluid (CSF) and brain

- Study the time window of drug/intervention efficacy

- Establish a correlation between the window of opportunity in animals versus humans

- Establish dose-response curve

- Obtain as much toxicological data as possible

Focus the trial on the appropriate subgroup of patients

- Target the subpopulation of patients most likely to benefit from the treatment

- Consider important prognostic variables: severity of injury, age, sex, hypotension, pupillary reaction, evidence of brain injury on computed tomography (CT), and traumatic subarachnoid hemorrhage

- Exclude patients who are "too good" or "too bad" to be helped by the test therapy

Confirm adequate drug delivery to the brain

- Establish the pharmacokinetics of the drug in humans

- Establish CSF penetration

Clinical management

- Standardize clinical management as carefully as possible within and across centers

- Conduct frequent independent monitoring of patient management and data quality

- Reduce the amount of data collection to only what is relevant

- Use centralized reading of CT scans

- Consider that concurrent medications may have significant interactions with the study drug

Choose the right outcome measures

- Consider carefully the point at which to dichotomize the Glasgow Outcome Score (GOS)

- Ensure that the time frame for outcome is reasonable; a 6-month time point is recommended for most trials

- Select outcomes that are measurable, standardized, and relevant to lifestyle

Surrogate outcome measures

- Do not accept a surrogate for clinical outcome in phase III TBI trials

- Use surrogate outcomes as supplementary data to guide future studies

- Consider surrogates as outcomes for phase II trials

Statistical considerations

- Power the study adequately

- Employ alternative designs to standard randomized clinical trial (RCT) where possible

- Have reasonable expectations for size effects; effect sizes of 5-7.5% could be valuable

- Consider large, simple trials to detect small treatment effects.

Informed consent

- Make every effort to obtain consent from the family whenever feasible

- Prepare an abbreviated consent form for the initial discussion

- Put into place appropriate mechanisms for waived consent before starting the trial

- Work closely with Institutional Review Boards (IRBs) to ensure communication with community and participants during study

Study management

- Establish stopping rules (safety, futility) in design of study

- Assure availability of drug/placebo formulations before entering patients

- Report study results within a year of completion of last follow-up, even when results are negative

- Establish independent safety monitoring board to oversee trial

Overview of FDA perspective

- A drug can be targeted either to a very specific subgroup of patients, or to the larger population—this is not prescribed by the law

- The Food and Drug Administration (FDA) does not specify the size of the treatment effect—the investigators have to decide what is clinically a worthwhile effect

- Placebo controlled trials would be the most appropriate for TBI since there is currently no approved drug

- The FDA ordinarily requires two independent positive trials for approval

- TBI is an appropriate candidate disease for fast track approval if a big effect is shown

- An effect on surrogate markers only, will probably not be acceptable for final drug approval

- Clinical outcome needs to be positively impacted

- International trials may be accepted if they are properly conducted, the patient populations are similar and the clinical management of patients is comparable.

LESSONS LEARNED FROM PREVIOUS TRIALS

Pharmacological Trials

Selfotel *(Ross Bullock, M.D., Ph.D.).* Selfotel was commercially developed by CIBA-GEIGY about 12 years ago, and it was the first glutamate antagonist to go into phase III trials. The drug went through a very extensive preclinical evaluation process, the kind of state-of-the-art drug evaluation process that a large company can bring to bear on a new compound. An important factor about this compound is that it is a competitive glutamate antagonist, and it binds to the same receptor site as glutamate. Most of Selfotel's early evaluation was done using preinjury dosing paradigms in animal models. There were five animal studies related to TBI, which were well done and some of them are published.

When phase I volunteer studies were done, it was found that this compound had psychomimetic/psychoactive behavioral effects. This side effect limited testing of the compound particularly in alert, awake patients. For TBI higher doses were used, because the patients were in coma. The drug went on to later Phase II studies, through The American Brain Injury Consortium (ABIC). This was the first trial of Selfotel in the United States.

CIBA was feeling pressure in 1992-93 to get the compound into the marketplace, and consequently full analysis of the data from the phase II study was not completed before the phase III study began. They launched four, large, "state-of-the-art" phase III trials, with a tremendous amount of input from experts in academia. They aimed to enroll about 1,200 patients into each of two stroke trials and about 860 into two trials of severe TBI. All four trials were negative.

What did we learn from this experience? Some of the animal studies had shown spectacular neuroprotection. There had been a few negative studies in animal models, but the balance of information strongly favored an effect. In one phase II study (108 patients) there appeared to be an intracranial pressure (ICP)-lowering effect, exactly what you would want to see with this kind of compound. At the time that the phase III protocol was being designed, we obtained data on the amount of glutamate in tissue from microdialysis. Because of the long duration of glutamate exposure in the tissue, at least four days of treatment with the drug was proposed in human TBI.

Four concomitant trials in stroke and TBI were in progress, and the stroke trials began to show higher mortality rates in the Selfotel-treated group. CIBA closed all four trials. With further analyses, the excess mortality rates disappeared in the stroke trials. It is possible that CIBA shut down the TBI trials on the basis of incomplete information. CIBA then did a futility analysis, after unblinding all of the data, and found no efficacy for the drug in any of the trauma subgroups. What did emerge was a lower overall mortality rate for TBI, in a well-monitored trial with good data quality.

What can we say in retrospect about the Selfotel experience? Adequate brain pharmacokinetics was never done prior to the large phase III studies, during early clinical work-up, or in the animal studies. Adequate drug binding to the receptor in the presence of high glutamate concentrations was never shown, and the drug was not measured in brain. We know that in stroke and TBI, glutamate is present in very high concentrations. Perhaps the drug failed to bind the receptor competitively in this environment. In the future, pharmacokinetics should be done early and thoroughly; however, it is possible that this compound could make a difference in TBI or in high-risk cerebrovascular surgery. It appears a safe and effective drug, in pretreatment paradigms that had been done in animal models.

Cerestat (Ross Bullock, M.D.). Cerestat (CNS 1102) was the leading product of a small biotech company, Cambridge Neuroscience (CNS). Much of the data regarding this TBI trial is still confidential, and has never been released or published. Cerestat is a noncompetitive glutamate antagonist. It binds in the channel site of the glutamate receptor, that is, at the magnesium binding site, and only binds when the receptor is activated by high concentrations of glutamate, so-called use-dependency. Thus, the drug should not bind significantly unless glutamate was increased in the tissue; the more channels that were open, the more drug would bind, and thus "damp down the ionic storm." The small size of the company and the relative lack of resources that could be put into development of the compound seemed to limit the amount of pre-clinical testing that was done.

The phase III protocol was designed with input from academia, including the European Brain Injury Consortium (EBIC); however, there were some points on which the sponsor had the last word. A 3-month GOS, which was at that time a departure from the norm, was the primary outcome measure. CNS enrolled 70 centers across Europe and the United States. Looking back at the data, over half of those centers enrolled fewer than five patients, a major factor in the outcome of the trial. A planned interim analysis at about 340 patients showed no benefit and no harmful effect. At the same time, the analysis of their stroke studies indicated lack of efficacy. To my knowledge, the final data analysis has not been published or presented. What can we learn from this trial? Large intercenter variability was probably enough to substantially degrade the quality of data.

CP 101-606 (Ross Bullock, M.D.). The most recently completed trial with a glutamate antagonist was by Pfizer with CP 101-606. A lot of information on this compound is currently under confidentiality, since there is ongoing data analysis. This compound was synthesized using techniques of molecular pharmacology and drug design. Molecular techniques were used to discover and clone receptors to which the compound could then be targeted. This is a "second generation" NMDA antagonist, and has fewer side effects than either Selfotel or Cerestat. Because the compound targets a specific subtype receptor (NR2B), it tends to have much cleaner pharmacology. The compound did not have any of the serotonergic effects that its predecessor, Eliprodil, had. The most interesting aspect of the compound is that it gets into brain tissue with apparently a fourfold higher concentration than seen in plasma. It achieves therapeutic concentrations quickly, and clears quickly when the infusion stops. Even at high concentrations, it produced no behavioral side effects in humans. At MCV, we used the compound in "open-label" phase II studies, and saw a potential beneficial effect and no bad side effects. The "phase IIb" trial is now completed, with 400 patients enrolled, and data analysis has been finished but not released. This seems to have been the best TBI trial to date in terms of protocol design, and development of the compound.

The mechanism of glutamate neurotoxicity is not disputed in neuroscience. These three compounds aimed at that mechanism have been highly effective in animal models, but have failed, in two of the three human trials. This is one of the biggest paradoxes that TBI researchers face. We must change the perceptions of industry and academia, if we want to persuade them to stay with this mechanism. Information from basic science presents many new fields to explore; however, in apoptosis, for example, there are no clinical drugs, the mechanisms are far less certain, and the contribution of apoptosis to TBI is not known and could be insignificant.

D-CPP-ene (Graham Teasdale, M.D.). The protocol for the Sandoz-sponsored study of the glutamate antagonist D-CPP-ene, is presented on the *Lancet* website. The data have been analyzed and reported in meetings. DCPP-ene was given twice a day for 5 days, and the recruitment time window was 12 h. The initial recruitment goal was 800 patients, but on statistical grounds the recruitment target was increased. The trial was completed when the new target was met, and 920 patients were recruited in less than 2 years in about 51 European

centers. The population was well balanced for early severity and CT scan categories. The protocol did not allow inclusion into the study unless there was clear clinical evidence of brain damage on the CT scan. Only eight patients were lost to follow-up, four from each study group. Overall, the patients who received the active drug had a slightly worse outcome at 6 months than the placebo group. The difference was not statistically significant. This result, along with findings in other smaller studies that were not completed, reduced interest in development of either this agent, or other glutamate antagonists for treatment of severe head injury.

When a lack of benefit is found, it is necessary to ask if the dose of drug was appropriate. There was evidence that subjects received enough drug to affect brain function. Whether it was enough to affect the damage might be debated. Early in the study, we became aware that some patients showed abnormal involuntary movements of a choreoarthrotoid type, associated with hypertension. It was very likely that they were drug-related, and to manage the events would result in loss of blinding. There was a protocol change so that sedation and paralysis were continued for 12 h after the last study dosage. The incidence of abnormal movements then fell quite considerably. Drug treated patients took longer to come off the ventilator, longer to recover motor responsiveness, and longer to leave intensive care. Another question is if the timing of initiation of treatment appropriate? Only 4% in either placebo or treatment group began treatment within the first 4 h after injury. For those treated in the first 4 h, outcome was better in patients who received active drug. In comparing results across studies, there is a trend over time for fewer patients to receive treatment in that first 4-h "window."

Question: You said there were well-done animal studies that had never been published. Why aren't they in the public domain? Why will investigators sign on to do such studies when they will not be published?

Answer: The majority are negative studies, and it is difficult to have a negative study published. Most such studies are contractual and reimbursed by the companies.

Question: Why do you make so much of the fact that Cerestat is a noncompetitive antagonist?

Answer: It is a critical issue pharmacologically. This characteristic confers important benefit to the compound for use in TBI, where glutamate is markedly elevated. We did not have enough information from pharmacologists in the beginning with Selfotel; for Cerestat, we did.

Question: Do you think there was any possibility of detecting a treatment effect in the Selfotel trial based on how the trial was done?

Answer: In animal studies, with pretreatment, Selfotel worked extremely well. But, human TBI trials are not a pretreatment paradigm. We rationalized the design on the basis that "40% of TBI patients deteriorate later, so we will focus on those people." However, we did not see evidence of subgroup efficacy in those patients.

Comment: Graham, I would like to congratulate you on what I think is a good use of subgroup analysis. It is certainly legitimate to take a subgroup analyses and generate from that a hypothesis that goes into another trial; indeed, it would be a mistake for the trialists not to do that analysis.

Answer: The issue of starting treatment soon after injury was touched on earlier. This may be possible with an agent that either has been in widespread use, or there is clear evidence of its safety. Treatment could be started without prospective consent, with administration less than 4 h from injury. I do not think that we can expect to start treatment so quickly unless we are using waived consent.

Question: You provided some plasma levels for D-CPP-ene and said that you were convinced that this was an active brain plasma level. How do you know in a disease as complex as an injury what a sufficient plasma level is?

Comment: The plasma levels achieved in patients were comparable to the plasma levels achieved in animal models where efficacy was shown; also, abnormal movements were seen in patients, and a higher dose would have been inadvisable.

Question: Over the course of a trial, how does an investigator ensure that the centers adhere to the protocol? How do we make the variability of management as small as possible, so that it is not a major confounding factor for outcome?

Answer: The high protocol compliance in the EBIC study of D-CPP-ene was illustrated in the very small number of protocol violations that led to exclusion of a patient's data. Of 900 patients, only five cases were protocol violations, there was 99% follow-up. I think that there is an "overconcern" about variations in management and how this might influence the results of a trial. Randomization within each center is a simple way to minimize the issue. Moreover, variations in management could be an advantage in enhancing the validity of the result. At an early stage, looking for proof of concept, it may be helpful to have a very uniform population, treated in "high-compliance centers." In contrast, in the definitive study aimed to determine the merit of a treatment for general use, management variations must be accepted.

Steroids *(Raj Narayan, M.D.).* Steroids have been in common use in neurosurgery since the 1960s and were initially used to treat brain edema associated with brain tumors. The effect was usually dramatic, and there was no doubt that glucocorticoids had an important role in the management of these patients. Laboratory studies showed that steroids reduced free radical production and had a protective effect on the brain. Consequently, steroids became commonly used in the management of a variety of neurological conditions, including head injury. In 1976, Gobiet compared low-dose and high-dose dexamethasone in 93 severe head injury patients to a previous control group and reported a benefit in the high-dose group. In the same year, Faupel reported a favorable response on mortality using steroids in a prospective, double-blind trial of 95 patients; however, there was a significant increase in the number of vegetative survivors and no improvement in favorable outcome.

There were subsequently at least six studies of steroids in TBI that showed no clear beneficial effect on outcome or ICP. Gaab reported no significant improvement in a prospective, randomized, controlled trial of ultra high dose dexamethasone starting within 3 h of injury. Grumme reported a prospective, controlled, randomized, multicenter trial of 396 patients with the steroid triamcinalone. There was a trend towards better outcomes in steroid-treated patients, especially if they had a GCS of <8 and had a focal lesion on CT scan. This combination was seen in 93 patients; however, there was no statistically significant difference between the treatment and the placebo groups at discharge or at 1 year. The authors concluded that treatment with steroids was potentially helpful in a subgroup of patients with TBI.

Two additional studies related to steroids are presented in more detail later. Marshall reported on the results of a large trial of tirilazad mesylate in TBI. This drug was believed to be more potent than traditional steroid formulations, without the glucocorticoid side effects. No overall benefit on outcome was detected. In a meta-analysis of randomized controlled trials in acute TBI, Anderson reported no clear benefit, but stated that there was a possibility of a small effect. As a result, the group recommended that a larger trial of over 20,000 patients be conducted in order detect this effect.

In summary, the data available to date does not demonstrate a clear beneficial effect of steroids in severe acute TBI. It has been speculated that certain subgroups may benefit; however, this

is not proven. The evidence-based *Guidelines for the Management of Severe Traumatic Brain Injury* state, "The use of steroids is not recommended for improving outcome or reducing ICP in patients with severe head injury."

Tirilazad *(Lawrence Marshall, M.D.).* The Tirilazad database represents approximately 1,700 severe head injuries, internationally and from the United States. I am going to make a few comments about those studies: what was right and what was wrong. I am not going to comment on the pharmacology, because there are complex issues in the science of free radical scavenging that would require much more time. Country-specific differences, or differences in care, are a significant issue: the difference in mortality for contusions was dramatic between countries, and in the United States between centers. Intercenter variation, not explicable by GCS or CT scan, was responsible for 40% of the variation in the Tirilazad trial. Some differences could be explained by overhydration initially. Patients appeared "overresuscitated"—blood pressures were elevated, as were ICPs and treatment intensity levels (TILs). In patients with intracerebral hemorrhages, the ones who came in with lower GCS scores, higher lesion volumes, and were operated on early had better out-come than those patients who had smaller lesions initially, a higher GCS, but who then deteriorated. The mortality was half for the patients who were operated on early. Such regional differences in care are important when looking at results overall, and serve to emphasize the need to have a high level of protocol adherence. There is another lesson here. We need to think through what the target should be in our clinical trials. Results from the Tirilazad trial indicate that high ICP—whatever the cause of the elevation—needs to be a primary target.

An additional finding from the Tirilazad trial concerned the wide variety of patients included, and the necessity to assure that the treatment groups were well balanced. We had real imbalances in both the Selfotel and Tirilazad trials concerning frequencies of CT scans and the variety of different CT diagnoses that were made. We know that these areas are associated with different patterns of outcome. For example, the presence of traumatic SAH skews outcome. In the Tirilazad trial and in the Selfotel trial, we found that patients who had minimal intracranial pathology had an extremely low mortality (<5%), and almost 80% had a favorable outcome. The inclusion of many such patients is likely to "front-load" a trial, and make it more difficult to find efficacy of a treatment. In contrast, patients who have one lesion greater than 5 cc have less favorable outcome, with a mortality rate more than three times that of the patients with minimal pathology. How do we classify the patients? For the most part, we have used stratification of the GCS, but the CT scan correlates better with outcome, except at the very bottom of the scale. If we do use CT scans, then there has to be a centralized reader. In our experience in the United States, almost a third of the scans had been misinterpreted or miscoded using the Traumatic Coma Data Bank classification.

One positive, but still troubling, observation from the Selfotel trial was the extraordinary performance of patients in the placebo group on the GOS. Clinical care has improved dramatically over the last two decades. Are we fighting ourselves by sticking to a certain assessment of outcome? It is going to be difficult to do better than a mortality of 7% in patients with epidural hematomas, as occurred in the Selfotel trial. Patients in the placebo group are doing very well with diagnoses of subdural hematoma. A few years ago, we would have expected that they would do poorly and a new treatment would have room to move people up into better outcome categories. We need to measure outcome better than we presently do.

Another variable that needs further discussion is traumatic subarachnoid hemorrhage (SAH). Using the grading system developed by Gabrielle Morris for SAH, we have seen that mortality varies from 13% to 44% among patients. Recording subarachnoid hemorrhage as "absent or present" is inadequate, and we more need data to help us develop a better classification of traumatic SAH.

To highlight the importance of balance within a trial, we need look no further than the experience with female patients in the Tirilazad trial. There was such a marked imbalance in the frequency of shock (4:1 for the Tirilazad group) that it masked our ability to analyze the data. Thus, given the relatively small number of women in the trial (incidence of severe TBI is lower in women than men), the data are meaningless. An additional observation from the Tirilazad trial, which also appears to be true in the Selfotel trial, is that the mortality rate for women is higher; but, in women who survive, cognitive outcome is better. We need to pay more attention to differences between men and women when we look at outcomes.

In our analysis of the Selfotel trial, we noted that the initial value of the ICP is a much more important predictor of outcome than cerebral perfusion pressure (CPP), as long as the CPP is at least 60 mm Hg. Neurological deterioration, which occurred in approximately 30% of patients, was usually associated with asymmetry of the pupils and increased the likelihood of death almost six-fold. These patients would likely benefit from a new treatment, and we should begin to concentrate our efforts in that select population. But, we must identify the patients early in the course of the injury. These results also suggest that treatments aimed at improving intracranial hypertension would be appropriate.

In summary, based on our previous experience with two large international trials and one large U.S. trial, we would recommend that one should exclude patients in the diffuse II category with lesions that are less than 5 cc in size. This criterion would lead somewhat surprisingly to a more equal distribution of patients with shock among treatment groups. Trial design should also take into account targeting elevated ICP: it kills 75% of the patients who die in the first week and contributes to many later deaths.

Question: When you say that we should make ICP the target, do you mean that it should be the primary outcome in clinical trials, or we should look into developing trials to treat intracranial pressure?

Answer: We suggested in the paper on the results of neurological worsening that it could certainly be a surrogate endpoint. It needs to be validated against a harder endpoint. I would say that if you do not see a change in ICP within a trial, you are not going to be able to show efficacy. These are the patients who ultimately go on to either have a poor outcome or die.

Comment: Another message from your study would be that one needs to see that the treatment lowers the initial ICP and the occurrence of worsening is lower in the treatment group; not that the treatment group does better for a given ICP. You can have a drug in which the treatment group had a lower ICP, but no better outcome, as was found in the phase II Selfotel studies. How soon was ICP measured in your study?

Answer: In the European Tirilazad trial, ICP was measured when the monitor was first put in, an average of 5.1 h from the time of injury. In fact, 70% of patients had it within 3.5 h. There is a trend in patients without surgical lesions that early aggressive attempts to reduce ICP favorably influences outcome. When patients went to the operating room earlier, the 6-month outcome in patients with intracerebral hemorrhage was dramatically better than in patients with delayed evacuation of hematomas—even though those patients were initially better clinically. To me, that is a hint that the ICP is critical, particularly in patients with an extraaxial lesion, but also in those with diffuse brain swelling. If one could treat high ICP earlier and more effectively, it should pay off both in terms of neurological deterioration, and with 6-month outcome.

Question: Your point is well taken that the assessment of the patient CT scans should be centralized; however, trials are getting larger and larger. Should they require central control and assessment of CT scans for the Corticosteroids Randomized After Significant Head Injury

(CRASH) study with 20,000 patients? Also, you suggested several times that subarachnoid hemorrhage should be quantified. How can that be done in a huge trial?

Answer: Remember that CRASH focuses primarily on minor TBI, and only to some extent includes moderate and severe injuries. I think that for milder injuries, CT abnormalities may not need to play a role in the study. In severe injury, you can have variations in mortality rates of 6-10%, based on the CT diagnosis, if you have a large number of patients in the DI I or DI II categories. Either we limit a study to a targeted group of TBI patients, (remembering that the scan results change in 30% of the patients), or we include the entire spectrum. If we include all ranges of TBI, we need to make certain the treatment and control groups match.

PEG-SOD *(Byron Young, M.D.).* I am going to discuss the results of the PEG-Orgatine (PEG-SOD; superoxide dismutase) study. As you know, free radicals contribute to the generation of secondary injury. Before this study, there were a number of experimental models and clinical trials suggesting that free radical scavengers would improve the outcome from severe head injury. Probably Kontos and colleagues did the most important animal work in the early 1980s, and later Paul Muizelaar reported results of a phase II trial that suggested a benefit in head injury patients. Based on these and other studies a randomized, parallel, placebo-controlled, blinded, multicenter trial was done. The hypothesis was that PEG-Orgatine, which was a free radical scavenger, would prevent secondary injury and, therefore, improve the outcome from severe head injury. The trial was done in 29 centers in the United States, and there were 463 severely head-injured patients in the study. The patients received an intravenous dose of either placebo, 10,000 units, or 20,000 units of PEG-Orgatine within 8 h of injury. The primary endpoint was GOS at 3 months. Patients with GCS 3 were included in the trial, although some secondary analyses that eliminated results from these patients were done later. The planned secondary endpoints were mortality and the Disability Rating Scale (DRS). The distribution of patients receiving drug and placebo was as it should be. The bottom line of this trial was that there was no significant difference in neurological outcome (GOS, DRS) or mortality between the patients treated with PEG-Orgatine and those receiving placebo.

There were, however, better but not statistically significant outcomes in the patients who received 10,000 units/kg PEG-Orgatine compared to those who received the placebo or the 20,000 units/kg dose. At 3 months, there was an absolute difference of 7.9% improvement, and at 6 months, a 6% improvement using the dichotomized GOS (good recovery or moderately disabled vs. severely disabled, vegetative or dead).

What questions were raised by this study, and what have we learned? Why weren't the statistics significant even though there was a measurable difference? Was there is a type 2 error? The study was designed to detect a difference of 14% with a 90% power. Was that adequate? The trial size and the treatment differences sought were based on a phase II trial. Did we miss a clinically significant difference between two arms of treatment? We would a new trial to detect a smaller, although important, treatment effect. What about standard care? There was good adherence to the protocol, but at the time we were not focused on the intercenter variations in routine treatment of TBI. That may be very important. The only statistically significant difference in this trial was that the patients treated with 10,000 units of PEG-Orgatine had a decreased incidence of acute respiratory distress syndrome (ARDS), but that did not seem to make a difference in mortality.

IGF-1/growth hormone *(Byron Young, M.D.).* We have just completed a single-institution, controlled trial in which we studied whether or not the administration of insulin-like growth factor (IGF)-1 and growth hormone and would improve neurological outcome and alter metabolic sequelae after severe head injury. IGF-1, which is produced in the liver, mediates the effects of growth hormone, and is important for organ growth. It stimulates glucose uptake,

glycogen synthesis, net protein synthesis, amino acid transport, and DNA synthesis; and, it causes cellular proliferation both systemically and within the central nervous system. Patients who have a severe head injury are hypermetabolic. Their actual metabolic rate is about 40% more than their calculated metabolic rate, and they are in negative nitrogen balance. You can provide enough nutrition to give them a caloric positive balance, but nitrogen repletion alone will not result in positive nitrogen balance. By administering growth hormone (GH) and IGF-1 we wanted to push these patients to a positive nitrogen balance. We thought this might improve clinical outcome by reducing or preventing anabolism, and reducing infection, preventing secondary injury, and helping the reorganization of the central nervous system.

The particulars of the study were as follows. The primary endpoint was improvement in nitrogen retention. We also wanted to determine whether this could be done safely: in the literature there are reports of adverse effects, primarily from lowering glucose. The patients in the study had GCS scores 4-10. We started the IGF-1/GH within 72 h of injury, and we used traditional recommendations for nutritional support. There was no difference in the demographics between the control and IGF/GH groups.

We were able to demonstrate that we could achieve sustained positive nitrogen balance in patients with severe head injury; however, this made no difference in the outcomes measured. There is no difference in the DRS. We also did quite a battery of neuropsychological tests, and there were no significant difference between groups at any time throughout the trial. We followed these patients for 24 months. There were more infections in the patients who were treated than in the control group (not significant). The conclusion of this study is that IGF1/GH blunts the metabolic sequelae of TBI, but does not improve the neurological outcome, and may increase the risk of infection. After we finished the study, we compared the pulmonary complications in our patients with the ABIC database and it appeared that the patients that received IGF-1 and GH showed a reduction in the incidence of ARDS.

At the start of the trial Genentech provided the IGF-1. In the middle of the trial there was a corporate restructuring, and they decided to no longer supply the IGF-1. After about 6 months of conversation with our university lawyers, Genentech reconsidered. About 6 months after that, the European trials with GH and IGF-1 were stopped for safety reasons. Even though we saw no safety problems, our study was also stopped. Did the early end of the trial affect its outcome?

Comment: We have a huge discrepancy among TBI trials. On one hand is the small PEG-SOD study showing a 60% better outcome at 6 months, possibly an 80% better outcome at an earlier time point, for a drug that is given in one dose and is completely safe. On the other hand, sits the CRASH trial, designed to include 10,000 patients treated with high-dose steroids, for which no efficacy has been shown before and where there are well-known side effects. Surely PEG-SOD is the better drug for a mega-trial. Focus such a trial on the emergency room with a single dose given as early as possible, when free radicals are known to be active.

Comment: There has been expectation that one study will answer everything—a so-called mega-study. Isn't this a bit of an illusion? There is a strong desire to examine a trial that is negative on its primary, protocol-specified outcome measure, in order to find a positive "subgroup." People are looking at subgroups retrospectively and saying, "Well, here is where the action is. Let's do a large trial looking at this." It is treacherous to do that.

Nimodipine *(Graham Teasdale, M.D.).* Studies of Nimodipine in head injury began in 1987. At that time, there was a low level of interest in pharmacological treatment of TBI, apart from trials of steroids with varying doses in smallish numbers of patients. There was clear evidence that Nimodipine was beneficial in spontaneous subarachnoid hemorrhage, reducing the incidence of infarction and ischemia and improving outcome.

Two studies were done in relatively unselected head injury populations: HIT-1, involving five British centers and Helsinki, and then HIT-2, with about 12 centers in different parts of Europe. These studies showed a 4% absolute improvement and 8% relative improvement for favorable outcome, which was not significant. Both of these involved a 7-day treatment with Nimodipine. In HIT-1, 20% of patients were recruited and treatment started within 4 h of injury. HIT-2 was started while HIT-1 was still going on; however, in this second trial, the proportion of subjects entered in less than 4 h fell to 10% of those recruited, and the relative difference for favorable outcome dropped to 2%. Although these results were disappointing, the recognition of the drug's effect in spontaneous subarachnoid hemorrhage, along with evidence from the Traumatic Coma Databank on the importance of SAH in head injury, prompted examination of the HIT-2 data to see what was going on. This analysis showed overall greater mortality in patients graded as having traumatic SAH on the CT scan, but outcome was better (8% improvement) in patients treated with Nimodipine.

This finding stimulated two activities: a further retrospective analysis of the HIT-1 database and a new prospective study. The study of the HIT-1 database indicated that traumatic SAH patients did slightly worse with Nimodipine than placebo. In contrast, a clinical study (HIT-3) involving 123 patients with CT scan evidence of SAH, admitted to 12 centers in Germany, showed a difference in the rate of favorable outcome for Nimodipine-treated patients. In a meta-analysis of these studies, there was a "just-significant" improvement in mortality and disability in Nimodipine treated-patients. HIT-3 had some limitations. For example, on review of the 123 patients admitted, 20% were found not to have subarachnoid hemorrhage. Although the overview resulted in Nimodipine being registered for the treatment of traumatic SAH in some countries, it has not been widely accepted as a standard treatment. A dialogue between clinicians and Bayer prompted the company to do a further prospective study, HIT-4. Results are not available at the time of this report.

This sequence of studies illustrate how an interesting subgroup can emerge from trials, and be evaluated retrospectively and prospectively. In the end, through Bayer sticking with the area for 12-13 years, a definitive answer will be obtained.

Question: It is my understanding that before the Nimodipine trials, there was not a single published preclinical study in any animal model of head injury. I was wondering how the company could possibly have convinced themselves to launch a large-scale clinical trial with no preclinical data available?

Answer: It was firmly established in the 1970s that ischemia is a major component of brain damage after head injury. Demonstration that Nimodipine had a good effect in models of ischemia seemed to be sufficient to move to patient studies. At that stage showing a good effect in a model of ischemia seemed to be sufficient to jump to patients with either spontaneous SAH or head injury. It was not just the company that was convinced, but the investigators as well. We were correct for spontaneous bleeding; the position in head injury remains to be defined.

Bradycor (Anthony Marmarou, Ph.D.). There had been a number of animal studies using bradykinin antagonists that dealt predominantly with brain swelling. It was on the basis of these preclinical studies that SmithKline Beecham elected to use intracranial pressure as the primary endpoint in the Bradycor trial. A prospective, randomized trial was designed, and began in 39 centers in the United States with coordination by the American Brain Injury Consortium (ABIC). After 139 patients were accrued, the company placed the trial on clinical hold. The hold was based on new animal work using a second batch of the drug conducted during the course of the trial. The first batch of the drug was made in the United States, and this was the one that we were using in the trial. The second batch was formulated in Europe, and samples from this batch were tested in rats by a contract laboratory. In a study with 12

animals, all of the rats died. This event was startling to us because we had observed no safety concerns with our patients in the trial, and our preclinical data had been clean. We (ABIC and the clinical centers) continued to hold, but presumed that this issue would be resolved by additional studies; however, the trial was stopped and the blinded randomization was broken. Therefore, I can only report the data that we had available at that time; the numbers will be small in the various groups.

There were no significant differences between control and treatment groups in any of the adverse event categories. The study groups were balanced with regard to the intracranial hypertension, cerebral hemorrhage, and in all other factors.. With regard to the primary endpoint (the percent time ICP. 20 mm Hg), there was a trend for reduction of ICP with Bradycor, although not significant. Looking at the entire intent-to-treat population, there was certainly variability in ICP, but a trend toward better control with treatment. With regard to outcome, we saw a reduction in mortality in the drug group. Analyzing the dichotomized GOS at 3 months, we saw close to a 10% improvement. This trend was seen at 6 months with a difference in favorable outcome between placebo and drug, but again not significant considering the small number of patients. Interestingly, we observed an improvement (not significant) in neuropsychological indices in the drug-treated group at 3 and 6 months. It is important to note the cooperation by SmithKline Beecham in assisting us in proceeding toward publication of a so-called negative trial.

What have we learned? Firstly, considering the favorable effect upon GOS, it is possible to include patients with GCS 3 and one reactive pupil. This trial shows that, even though there is devastating injury as evidenced by the low GCS group with one reactive pupil, some patients do show measurable improvement. With regard to interim analysis, we believe that guidelines for breaking the blind should be firmly established. In this case, the 100% rodent mortality using the second batch of drug could not be reproduced. This was confirmed by a second contract laboratory. It was thought that the rats were hyperthermic as a result of a heating lamp used during the testing, but no further work was done to assess the true safety of the second batch of drug. The investigators involved in this trial had no voice in the decision to break the blind and stop the trial. We believe that we owe it to our patients, to complete a trial whenever possible, and to reach a scientific conclusion.

One lesson is to do as much as possible before the trial starts. Stopping rules (safety, futility) should be stated clearly and agreed to at the outset. The manufacturing process and drug supply should be in place prior to entering patients.

Dexanabinol *(Nachshon Knoller, M.D., and Anat Biegon, Ph.D.).* Dexanabinol is a novel synthetic chemical analog of the active component of marijuana. It was designed as a mirror image of the naturally occurring compound, so it is not recognized by the cannabinoid receptors in the brain that mediate the intoxicating effect of marijuana. Preclinical studies demonstrated that Dexanabinol is a noncompetitive inhibitor of the NMDA receptor, a free radical scavenger and antioxidant, and an inhibitor of the pro-inflammatory cytokine TNF alpha. Dexanabinol inhibits breakdown of the blood-brain barrier, edema formation, and neurological deficit in a closed head injury in rats with a therapeutic window of 6 h. It has also been shown to be efficacious in models of axonal crush injury to optic nerve, and in focal and global ischemia. Following completion of a phase I safety evaluation in healthy volunteers, work started on the design of the study protocol in severe head trauma. Several phase II and phase III trials of neuroprotective agents have been already completed or ongoing (PEG-SOD, hypothermia, Tirilazad) or were on hold (Selfotel) at that point in time, such that lessons learned from these trials could be incorporated into the design of the Dexanabinol protocol. Two major design principles were implemented based on the analysis of pre-clinical data and data from previous

trials so as to increase the likelihood of detecting a drug effect: (1) tighten inclusion/exclusion criteria and (2) follow the lead from preclinical studies.

Criteria were chosen to exclude patients with a high likelihood of extreme outcome (death or good recovery) regardless of treatment. To achieve this end, patients with enrollment GCS 4-8 were included unless both pupils were fully dilated and fixed. The latter subgroup was excluded due to extremely poor prognosis. Evidence of intracranial pathology was required (CT category 2 or above) but patients with pure epidural hematomas were excluded since this subgroup has an extremely good post-surgical prognosis. Finally, due to the potent antiedema effect of the drug in animal models, there was a strong emphasis on monitoring ICP (a measure of the brain edema) in this study such that only patients requiring ICP monitoring were enrolled.

Treatment parameters (time to treatment, dose, and duration) were kept as close as possible to those established as safe and efficacious in relevant preclinical models. Thus, Dexanabinol was delivered within 6 h of injury (window of efficacy established in closed head injury and axonal crush models). The doses were derived from comparative pharmacokinetics in animal studies where effective doses in rats (2-5 mg/kg) produced peak plasma levels of ~2-5 mg/mL, and in the phase I trial. Plasma levels in this range were obtained in volunteers and patients at doses of 48-200 mg/subject, so this was the dose range chosen for the patient study. The drug was administered only once, since repeated administration in the animal models did not appear to provide additional benefits. In fact, the best treatment paradigm in rats consisted of one injection within an hour of injury and a second injection after 6 h (additional treatments at 12 and 24 h, and 3, 7, or 10 days showed no further benefit). Since an hour after injury was not considered a realistic treatment time to expect within the framework of a clinical trial, a single injection within 6 h was chosen.

The study was not powered for efficacy; rather the expectation was to derive safety data and some indicators for efficacy through influence on surrogate markers such as ICP and early recovery. The phase II trial of Dexanabinol in severe head trauma took place in all six neurosurgical units in Israel between 1996 and 1999. The study was a multi-center, randomized, double-masked phase II b trial of a single dose of Dexanabinol intravenous solution with a placebo control. A total of 101 patients were randomized to receive a 48-, 150-, or 200-mg dose of Dexanabinol, or an appropriate volume of vehicle. The primary outcome (safety) measures were ICP, cardiovascular function, and clinical outcome/adverse events. Secondary outcome (safety) measures were dichotomized GOS, GOAT, and DRS.

The study patients had the demographic profile expected for TBI: mostly young (mean age, 30 years), mostly males (80%), victims of motor vehicle accidents (70%). There were no statistically significant differences in important risk factors (GCS, CT, age) between the drug and placebo groups. Dexanabinol was shown to be safe and well tolerated in the dose range tested, as expected from the preclinical and phase I data. The drug did not change the adverse event profile, such that the observed adverse events and their relative frequency were those characteristic of the severe TBI population. There were decreases in the incidence of fever, hypotension and mortality, although the differences did not reach statistical significance. Significant effects of the drug were seen on ICP. The drug appeared to prevent the increase of ICP over the first 2-3 days postinjury, such that mean ICP values (which were initially similar in the drug and placebo groups) rose above 15 mm Hg in the placebo group and remained consistently below 15 mm Hg in the Dexanabinol group. The percentage of time ICP was above 25 mm Hg was decreased in the Dexanabinol treated groups at all doses, and the effect was statistically significant from the second day. This stabilizing effect of Dexanabinol on ICP was achieved without lowering systolic blood pressure; conversely, the percentage of time systolic blood pressure fell below 90 mm Hg was reduced in the Dexanabinol-treated patients.

Examination of the distribution of patients among the various GOS categories over time results in several expected and some unexpected observations. The trend towards lower mortality and higher percentage of patients in the "favorable" outcome categories, observed in recent clinical trials, continues in this study, such that mortality is below 20% and a more than 65% of patients achieve favorable outcome without treatment. Thus, dichotomized GOS used in the traditional way produces a ceiling effect, which makes it hard to show a drug effect. Although, as expected, there were no significant effects of treatment on 6-month GOS; trends towards better outcome in the Dexanabinol-treated patients were stronger in the more severely injured subgroups (GCS 4-6, CT. 2) and at the earlier time points (1 and 3 months). In fact, the percentage of patients achieving good recovery after 1 month was significantly increased in the Dexanabinol-treated group. Statistically significant differences in favor of the drug were found on the GOAT. GOAT scores were persistently better in the Dexanabinol-treated patients throughout the follow-up period.

These lessons should promote discussion of the design of future phase III studies in relation to patient selection as well as outcome measures. Excluding patients without CT evidence of severe parenchymal damage, or those above GCS 8 might increase the likelihood of a significant effect of treatment. Alternatively, one may consider the inclusion of patients with GCS 3 and one fixed dilated pupil. In terms of outcome criteria, it appears that a 10% difference in favorable outcome on the dichotomized GOS should be replaced by another, more innovative statistical analysis, which better, reflects the reality of current outcome distribution in severe TBI. We should also consider other endpoints with clinical importance, such as ICP/CPP management, evolution in CT pathology, shorter hospitalization, shorter ICU stay, and early recovery. From the clinical perspective as well as in terms of cost effectiveness and reduction of suffering for patients and families, a drug that facilitates ICP management, shortens ICU and hospital stays, and promotes achievement of the good recovery category after 1 month instead of 6 months, is a very good drug indeed.

SNX-111 (J. Paul Muizelaar, M.D.). Good preclinical studies are essential for a good clinical trial. We need to understand the mechanism of action of the drug. There should be improved outcome in experimental animals. For most of the trials that were designed before the trial of Dexanabinol or SNX111, that was really not the case. What preclinical data existed about SNX111? An important finding in studies of stroke mechanisms was that this drug was effective if given 24 h after transient forebrain ischemia. This long "window of opportunity" was one of the attractive features of the drug. In addition the effect appeared long lasting, and was still present on day 28 after injury. So, before the TBI trial was started, we determined what the time window for the drug was. In preclinical work, we chose a dose of 4 mg/kg and gave that to animals 15 min before the injury, and 15 min, or 1 h, 2 h, 4 h, 6 h, or 10 h after injury. We looked, not at outcome, but at a very specific measure of metabolic (mitochondrial) function in the brain. We found that this drug works very well if given 2-4 h after the injury. Then we looked at mitochondrial function, efficiency and found again that the drug given later was actually more effective than when given right after the injury. Next, we looked at different doses, and these doses were then always given 4 h after injury. In addition to 4 mg/kg, we also looked at 0.5 mg/kg and 1.0 mg/kg; 1.0 mg/kg was still fairly effective, but 0.5 mg/kg was not. So on the basis of these data, and after testing pharmacokinetics in animals and in human volunteers, a dose was chosen for this particular drug. We also felt that this drug in this dose could be successful in TBI because the time window allows a later administration. One important finding from our preliminary experiments was that by targeting mitochondrial function, we were able to get a dose and time window using only 39 animals. For behavioral outcomes, we would have needed 114 experiments.

The trial was designed in 1995 with standard inclusion criteria for severe closed head injury. The only thing I feel we did a little bit differently was to use a decision tree published by Choi.

The decision analysis combined parameters for patient outcome. We used this particular decision tree to derive strata in the trial for predicting those patients likely to have a good outcome. We predicted good progress in patients with bilateral pupillary responses, under 40 years old (y/o), and GCS > 4; or a unilateral pupillary response and <30 y/o; or a motor score of 5 and <25 y/o. Other patients were considered in a poor prognostic stratum.

What have we learned from this trial? The trial was suspended when it was found that mortality in the patients receiving the drug was higher than in the patients who were receiving placebo, and the company has not yet released the full data. A summary of what we know follows. We entered 160 patients before the trial ended. The strata for expected good outcome and expected bad outcome were approximately equal in the treatment groups. There were 40 patients treated with placebo who were expected to have a poor outcome, and 40 treated with drug who were expected to have a poor outcome; the same figures for patients with an expected favorable outcome. The mortality for the SNX arm was almost 25%, while for the placebo arm it was 15%. It was very disappointing, because we thought that we finally had a drug that could be given a long time after the initial injury. At the start, we knew that the drug caused hypotension; so, the design stated that treatment was never given if patients did not have a normal CVP, and if dopamine and/or phenyl-epinephrine were not already hanging as an infusion.

From this trial, we can see what very aggressive fluid management will do. I think I never saw better preclinical data with any drug; but the animals were not as severely injured as many of our patients. Perhaps that contributed to the disappointing outcome.

Anticonvulsants *(Nancy Temkin, Ph.D.).* The University of Washington has conducted three acute trials in epilepsy after traumatic brain injury: Dilantin, 1983-1989, Valproate 1990-1997, and, currently, magnesium sulfate. Briefly, the Dilantin study was a randomized, double-blind, parallel group study; 404 patients were entered, treatment was initiated within 24 h, and 24% of the subjects were lost during the 2-year follow-up. The primary goal of the study was to see if Dilantin prevented epileptogenesis. We found that the drug prevented seizures in the first week after head trauma, but, even though treatment continued for 1 year, there was no effect on late seizures either during the treatment period or after treatment was stopped. We found that phenytoin had substantial medical and neurobehavioral side effects, especially early on, in the more severely injured cases. The second study used a similar design to see if Valproate had an antiepileptogenic effect. We examined two durations of Valproate therapy, 1 month and 6 months, with follow-up continued to 2 years. Of 400 patients entered, about 15% were lost to follow-up. Conclusions were that Valproate showed no benefit over phenytoin for early seizures, and neither drug prevented late seizures. There were essentially no adverse neuropsychological effects in the Valproate groups; however, there was a trend towards higher mortality, the cause of which we were never able to determine. Our current study examines magnesium sulfate: randomized, double-blind, parallel group, placebo-controlled, about 400 patients. The treatment is initiated within 8 h, and the duration of treatment is 5 days. We are looking at 6-month follow-up rather than 2 years, using a composite endpoint. In addition to antiepileptogenesis, the study will evaluate survival and, primarily, neurobehavioral outcome.

If you have two studies, one with 100 patients targeted to the mechanism of a drug, and another study with 200 patients, but in 100 patients the drug is not appropriate and has no effect, then you will have the same power for both of these studies. If more than half the patients who are unaffected, the effect on the power will be more extreme. As much as possible, you want to know what a drug is supposed to do, and how to identify patients for whom that drug might be useful. In our Dilantin and Valproate studies, we were looking at effect on seizures, so we used patients who were at high risk for developing seizures. In the magnesium sulfate study, the drug effect is likely very broad, so the inclusion criteria are much less stringent than in the two seizure studies.

Comment: There are a couple of issues surrounding use of anticonvulsants to keep in mind. The first is a pharmacokinetics issue. Those of us who were involved in the Tirilazad trials learned a very difficult lesson, and that is that drugs like phenytoin, and even phenobarbital, stimulate liver enzymes and enhance the metabolism of drugs like Tirilazad. The second issue has to do with the potential impact of depressant-type agents. Anticonvulsants are depressant agents and may affect brain plasticity and recovery after trauma. Animal work has shown that administration of stimulants during a critical phase after injury facilitates behavioral recovery, and administration of depressants blunts that recovery. Wholesale, continuous use of anticonvulsants in "neurocritical" patients, has to be looked at very carefully. These trials are very important.

Comment: One has to ask if a particular outcome is an appropriate measure of the function that you would like to improve. What is the clinical utility of a particular outcome? Another issue concerns use of composite scores. While they may be appropriate, one has to ask, "What is the clinical interpretation of composites? What does such outcome mean?"

Physiological Trials

Cerebral perfusion pressure (Claudia Robertson, M.D.). The purpose of this study was to determine if a management protocol could be developed to reduce the incidence of the most common causes of secondary insults, particularly hypotension and hypocapnia, after severe TBI. The first hypothesis was that the treatment protocol, coined the "CBF-targeted" protocol, would reduce the incidence of secondary ischemic insults. The second hypothesis was that if the CBF-targeted protocol successfully reduced ischemic insults, the incidence of refractory intracranial hypertension would be reduced, and the third hypothesis was that neurological outcome would be improved.

Two management protocols were compared in a prospective, randomized, single-institution clinical trial. This information is summarized in Table 1. The control protocol was called the "ICP-targeted protocol." This control protocol consisted of a traditional TBI management strategy, where the primary goal of the treatment was to reduce ICP. The treatment or CBF-targeted protocol was designed to improve cerebral perfusion and prevent secondary ischemic insults. There were four major differences in the two treatment protocols. The first difference was in the end-points for fluid administration. In the ICP-targeted group, maintenance fluids were given. In the CBF-targeted group, fluids were given to maintain a normal central venous pressure or pulmonary wedge pressure. The second difference was the goal for mean blood pressure, at least 70 mm Hg in the ICP-targeted group and at least 90 mm Hg in the CBF-targeted group. The third difference was the goal for cerebral perfusion pressure, at least 50 mm Hg in the ICP-targeted group and at least 70 mm Hg in the CBF-targeted group. The final difference was in the use of hyperventilation. Although hyperventilation was not used in any patient as a routine, the ICP-targeted protocol allowed use of hyperventilation as a treatment of intracranial hypertension. The CBF-targeted protocol did not use hyperventilation, because, although it lowers ICP, it does so at the expense of reducing cerebral perfusion.

The inclusion criteria were adults (\$15 years), who were admitted within 12 h of a severe TBI and who had a motor component of the GCS of 5 or less on admission. Exclusion criteria included only brain death, contraindication to placement of a jugular bulb catheter, or severe associated systemic injuries.

The primary outcome measure for the trial was the incidence of secondary insults, indicated by the occurrence of jugular venous desaturation. The secondary outcome measures were the incidence of refractory intracranial hypertension, and neurological outcome measured by the 6-month GOS and the DRS. The sample size of 182 was chosen to give an adequate power to detect a 50% reduction in the incidence of jugular venous desaturation. This sample size would

also detect a reduction in the incidence of intracranial hypertension from 27% to 12%, and an increase in the number of favorable outcomes (good recovery/moderate disability) from 35% to 54%. It was recognized that this number of patients would only detect a very large effect on neurological outcome; however, this is the improvement in outcome that has been observed in clinical series with this type of management protocol.

The patients were randomly assigned to treatment by time blocks. Each 4-month rotation of neurosurgeons through the ICU was divided into 2-month treatment blocks. The treatment protocol was randomly assigned for each of the 2-month treatment blocks, and all patients admitted to the hospital were treated by the assigned protocol for that 2-month block. In this manner, each physician group had experience with both treatment regimens, but the order of this experience was random.

The study required eight 4-month blocks of time to enroll the required number of patients. A total of 89 patients were enrolled in the ICP-targeted treatment months, and 100 patients were enrolled in the CBF-targeted months. Because this randomization scheme does not completely protect against bias, the patients who were admitted but not enrolled in the study during the time period of the study were compared to the study group. The only differences were those that would be expected because of the exclusion criteria for the study. The excluded patients had a lower initial GCS, a higher frequency of abnormal pupils, a higher frequency of gunshot wounds as the cause of the injury, and a higher frequency of prehospital hypoxia and hypotension. The characteristics of the patients enrolled in the two treatment groups were also compared. There were no significant differences in any measure of demographic characteristics or initial injury severity.

The CBF-targeted protocol, was very successful in preventing secondary ischemic insults. The number of patients who had one or more episodes of jugular venous desaturation was reduced from 51% to 30%, and the length of time that jugular venous saturation was decreased was significantly reduced. This improvement occurred primarily in secondary ischemic insults caused by hypotension and hypocapnia, which were the two aspects of treatment emphasized by the CBF-targeted protocol. The reduction in ischemic insults occurred with the CBF-targeted protocol throughout the entire study, that is, it was not physician-dependent and it did not depend on which protocol was assigned first to each physician group. In addition, the treatment protocol remained a significant predictor of the occurrence of secondary ischemic insults, even when logistic regression analysis was used to adjust for injury severity. In the final best-fit logistic model, the ICP-targeted protocol was associated with a 2.4-fold increased risk of an ischemic insult.

The CBF-targeted protocol, however, did not reduce the incidence of intracranial hypertension. The mean ICP, the duration of time that the ICP was elevated, the percent of patients with refractory intracranial hypertension, and the number of patients who died of intracranial hypertension were similar in the two treatment groups. The CBF-targeted protocol, did not significantly improve neurological outcome either. It must be remembered that the study was not powered to detect anything but a very large improvement in outcome. However, the trend was for a higher morality rate, and a lower percentage of patients with a favorable outcome in the CBF-targeted protocol. Because of this apparent paradox, where a marked reduction in secondary ischemic insults did not result in an improvement in neurological outcome, potential complications of the CBF-targeted protocol, which might have confounded the ultimate outcome, were examined.

Three complications were possible from the treatment required to maintain an elevated CPP. A higher blood pressure could cause delayed or recurrent intracranial hematomas. Higher fluid intake and prolonged use of pressor agents could result in pulmonary edema, and high doses

of pressors could be associated with renal failure. Any one of these complications could have offset any beneficial effect of reducing secondary insults. Two complications were easy to eliminate: there was only one case of acute renal failure, and there was no difference in the incidence of either recurrent or delayed intracranial hematoma between the two protocols. The incidence of ARDS, however, was five times higher in the CBF-targeted protocol group than in the ICP-targeted group (15% vs. 3%, respectively). The aspects of treatment that were related to the development of ARDS included the following: a greater intake of fluid, a more positive fluid balance, a higher central venous pressure and pulmonary wedge pressure, and more prolonged use of pressors (both dopamine and epinephrine). Finally, the outcome in the patients who developed ARDS was significantly worse. Seventy-one percent of the patients who developed ARDS remained vegetative or died of their injury by 6 months postinjury.

To summarize, the CBF-targeted treatment significantly reduced the incidence of secondary ischemic insults, from 51% to 30%, but this reduction in ischemia did not improve neurological outcome. There may be several explanations for these findings. First the sample size was not large enough to detect a realistic difference in outcome. Second, because jugular venous desaturation was treated in both groups when it did occur, this minimized any adverse effect on outcome. Finally, the beneficial effects of reducing secondary ischemic insults may have been offset by systemic complications, especially ARDS.

The lessons that can be learned from the experiences of this study include the following: (1) Management trials can be done in TBI patients. There are many management issues that exist in neurocritical care, and it is only through systematic controlled studies that issues such as this can be sorted out. (2) The current recommendation of keeping all TBI patients at a CPP of at least 70 mm Hg needs to be reconsidered in light of the significant complications that were observed in this study and the lack of benefit on overall outcome.

Question: Given your findings, how would you change your management?

Answer: This study suggests that we need to individualize treatment. It does not make sense to maintain all patients at an elevated CPP, when perhaps only a few patients really need that level. In doing so, all patients are put at risk of the complications, while only a few receive benefits. As a routine now, we maintain patients with a CPP of at least 60 mm Hg, which is adequate for most patients. However, if some measures, such regional CBF or $SjvO_2$ or brain tissue pO_2, suggest that perfusion is inadequate with a CPP of 60 mm Hg, then we try to raise CPP. This puts only patients with potential to benefit at risk of complications.

Comment: I think that CPP is a red herring, and we are trying to increase a "parameter" that does not really exist. We need to consider that pulmonary failure potentially exacerbates the brain swelling, and we need to rethink the whole notion of the CPP calculation. We need to protect the brain against ischemia, which you have argued very well. An alternative explanation for your data is that we need to direct therapy more quickly, vigorously, and intelligently against ICP. What we are really looking at early on is cytotoxic edema and not diffuse intravascular swelling. How would you go after that?

Answer: Your points are well taken, and I would add some strength to what you have said. We did not change the incidence of ischemia associated with intracranial hypertension. We did not design our treatment of ICP-related events any differently in the two treatment protocols. And the results could suggest that ICP-related ischemic events are much more important than those associated with hypotension or hypocapnia. How to treat is a much more difficult question. I presume that this type of edema is from the primary injury, which will require a different treatment strategy.

Comment: One point stares me in the face from your data: You have effectively treated secondary insults, which we have all been targeting for so many years, but the primary insult overshadows that treatment. How do we define the mechanisms for the primary insult?

Question: Some of us have advocated utilizing two different compounds in treating acute head injury. From your management standpoint has anybody ever thought about the possibility of combining a particular compound or therapeutic agent with a management style that could dictate what sort of patient should be a candidate for a particular treatment?

Answer: I think that the individualization of treatment for a particular patient's injury is a way to approach the problems that we are seeing. We used a very specific definition of ischemia; that is, a global ischemic insult sufficient to reduce the $SjvO_2$ below 50%. I think that is good evidence that the brain is being hypoperfused. Whether it is actually ischemic or not is another question.

Question: Your previous studies have shown that patients who have few desaturations do much better. Have you not shown here that you can potentially improve neurological outcome, but overall outcome depends on the systemic complications? Therefore, should you try to find a way to eliminate the systemic complications, rather than concluding that improving the perfusion did not improve neurological outcome?

Answer: We would agree that reducing secondary ischemic insults is a good thing to do. However, trying to maintain all patients at an artificially elevated CPP may not be the best way to approach this because of systemic complications associated with this preventative treatment. One alternative that can be applied right now is to only elevate CPP in those patients who really need a higher CPP to adequately perfuse their brain.

National acute brain injury study: hypothermia *(Guy L. Clifton, M.D.).* For the recent hypothermia trial, our primary question was: Does early induction of hypothermia to 33°C for 48 h improve outcome with low toxicity? It did not.

The trial was set to detect a 10% shift in dichotomized GOS. That is Good Recovery/Moderate Disability versus Severe Disability/Vegetative/Dead. The inclusion/exclusion criteria were as usual for most trials in TBI: no gunshot wounds, randomization within 6 h of injury, exclusion of major multiple trauma. There were 11 U.S. centers in the trial at different times. Four centers randomized a total of 33 patients before they were dropped from the trial for various reasons, and two other centers came into the study rather late; therefore, five centers randomized 88% of the patients. From the very start of the trial, we paid serious attention to early cooling and complication rates. Our goal was to reach a core body temperature of 33°C by 8 h after injury. Temperature and hypothermia-related complications were reported to the Performance/Safety Monitoring Board on a regular basis. We saw no significant demographic differences between the two treatment groups (hypothermia vs. normothermia), so our randomization was effective. Within the hypothermia group, eight patients did not reach target temperature. In the normothermia group, the mean temperature in the first 96 h was 37°C. The hypothermia patients were cooled to reach 33°C by 8.4 6 3h after injury. Randomization was at 4.3 h after injury. Patients remained hypothermic for 48 h and were re-warmed very slowly over 18-24 h. From pilot work we knew that this slow rewarming was very important.

Overall, there were differences in management of the two study groups for only two areas. One difference was that hypothermia patients received more fluids, with a mean fluid balance of 3 L positive over 96 h. Normothermia patients were about 1.6 L positive. That difference was significant. Hypothermia patients also received vasopressors for about 80% of their total hours in the first 96 h, and normothermia patients 69% of those first hours.

There was a concern expressed during peer review of the grant application that this was an unblinded trial, and everybody wanted it to work; the personnel who most wanted it to work would manage the hypothermia patients differently in some way. So we evaluated treatment intensity using a standard scoring system to quantify the number and intensity of interventions. It was the same in the hypothermia and normothermia groups. Despite the use of hypothermia, there was no evidence that one group was managed more vigorously. We targeted a CPP of . 70 mm Hg based on the best information available before the trial. In the first 96 h in ICU, 97% of the patients had a CPP under 70 mm Hg at some point, so virtually everybody experienced an episode for some period. About 40% of the patients had a CPP under 50 mm Hg at some point. The "percent hours under 50 mm Hg" was small (about 2% of hours); the "percent hours under 70 mm Hg" were about 10% of the hours monitored. So, while patients did drop under 50 mm Hg, they did not stay there long.

Our data show that there are critical management thresholds which adversely affect outcome if they occur at all: CPP, 60 mm Hg, ICP. 25 mm Hg, and MAP, 70 mm Hg. The incidences of these critical thresholds were the same in both treatment groups.

We saw a difference between the incidence (time) of ICP. 30 for the hypothermia and normothermia groups, and it was a big difference, about 25%. We found a decrease in the numbers of patients with ICP. 30 mm Hg and .40 mm Hg in the hypothermia group. It seems that hypothermia reduced the incidence of high ICP.

The mean arterial pressures (MAPs) were not different between the groups; but, considering the occurrence of MAP, 70 mm Hg, there were more of these patients in the hypothermia group. The MAP reduction was offset by the significant ICP reduction, so that if you look at the hourly occurrences of CPP, 50 mm Hg, you do not see a difference between the groups. Overall, there was a slight but statistically significant increase in complications in the hypothermia group. There was no treatment effect in terms of the primary hypothesis: 56.9% of patients had poor outcomes in the hypothermia group, 56% in the normothermia group. The mortality rate was 28% with hypothermia and 27% with normothermia.

We looked at the effect of hypothermia on GCS subgroups, age subgroups, presence of operative hematomas, subarachnoid hemorrhage, time to target temperature, and admission temperature. An effect of hypothermia was seen in older patients and in patients who were hypothermic on admission. We found that in the patients over 45 years old, there was an 89% poor outcome in the hypothermia group, and a 69% poor outcome in the normothermia group. This difference was highly significant. The cause of this outcome was an increase in medical complications; there were statistically significant increases in bleeding, sepsis, and pneumonia in the hypothermia group over 45 years of age.

Over 20% of patients were hypothermic (≤35°C) on admission, and the mean temperature over the first 8 h for this group was about 33.7°C. The mean admission temperature for the rest of the patients was about 36.2°C. The patients who came in hypothermic and who were randomized to normothermia warmed very slowly over 16-18 h after admission. Patients hypothermic on admission who were randomized to hypothermia had 61% poor outcome; "hypothermia-on-admission" patients randomized to normothermia had about a 78% poor outcome. This difference is significant.

The overall treatment effects for the five largest centers are summarized: Houston had a 5% positive (15%) effect for hypothermia; Sacramento, 24%; Pittsburgh, 114%; St. Louis, 214%; and Indianapolis, 220%. There were significant differences in CPP management among centers; however, we cannot detect a correlation with treatment effect. It appears that differences in baseline variables among centers in numbers of older patients and/or patients

who were hypothermic on admission caused this effect. My conclusion is that patients over 45 years old should not be cooled. In addition, I think we missed the treatment window.

Question: If you sort out the subset that was GCS 5-8, comparable to the phase II Pittsburgh study by Marion, did that stand up with a significant effect?

Answer: No. We found a 7% treatment effect in the GCS 5-8 patients; Marion reported a 38% treatment effect. In his study, he actively rewarmed patients who were hypothermic on admission over a period of approximately 6 h. In the laboratory, rapid re-warming worsens outcome. In our phase III study, we allowed patients to rewarm spontaneously. In Marion's phase II normothermia treatment group, there was a 66% poor outcome for the GCS 5-8 patients. Our phase III trial showed a 52% poor outcome for this group. There had been an unexpectedly bad result in normothermia patients the earlier Pittsburgh study. Dr. Marion and I think that the discrepancy may relate to active rewarming of the patients. In addition, I think the effect seen at Pittsburgh in our phase III trial may have been due to an increased number (36%) of patients who were hypothermic on admission.

Comment: In stroke patients who were rewarmed over 24 h the fatality rate was directly related to increased ICP. If patients are rewarmed over 48 h, there is none of the elevated ICP. Frankly, I think that a problem in your phase III trial was that the treatment groups were not well matched.

Answer: Fewer older patients in the hypothermia group in combination with admission hypothermia would maximize the probability of a treatment effect in any center.

Comment: Perhaps hypothermia should be viewed as a potential measure to buy time for the introduction of other therapies (e.g., drugs), rather than as a primary treatment.

SCIENTIFIC ISSUES

Developing New Pharmaceutical Agents (Edward Hall, Ph.D.)

How do we maximize our chances for success in developing effective drugs for head injury? First, there must be a target mechanism. We need to demonstrate the time course of that mechanism in relevant animal models. For any potential therapy, we must perform rigorous dose-response studies to see the effect of the drug (or treatment) on the target mechanism and any associated pathophysiology. For studies advocating neuroprotection, we actually need to demonstrate histological preservation. Ideally, we should show some behavioral benefit of treatment before going forward clinically. The neuroprotective action should correlate with plasma and brain pharmacokinetics. We have not done a very good job of this for many drugs considered in phase III trials.

We need to come up with relevant, easily available biomarkers. For instance, lipid peroxidation products have been shown to increase in the plasma in ischemic brain injury and in subarachnoid hemorrhage patients. We could use such biomarkers to correlate plasma and brain pharmacokinetics with modification of the marker in plasma and brain. This relationship would show that the drug could actually affect the selected target mechanism.

We need to compare single versus multiple dose regimens. Much of the preclinical evaluation of neuropharmacological agents involves administration of a single dose at some time point. Regimens studied have been totally empirical in the majority of preclinical and clinical trials. There are very few examples of systematic evaluations of the efficacy of a single dose versus sustained dosing; and, looking at sustained dosing, the duration of the treatment. An example is found in the National Acute Spinal Cord Injury Study (NASCIS) II and III studies, and the story of methylprednisolone in spinal cord injury. Based on careful preclinical studies, it was

apparent that bolus administration plus an infusion for some period of time made the most sense. I think that for many kinds of agents that we have considered for trauma, this type of regime would make sense. We simply have not done adequate therapeutic window studies in our animal models in most cases. Furthermore, when we do have such data, we tend to ignore it when we go to the clinic. Our excuse has been that the therapeutic window in a rat is probably not relevant to humans; however, we do not really know that.

I will quickly summarize how we evaluated methylprednisolone in acute spinal cord injury models, and how that preclinical work successfully translated into clinical efficacy. The mechanism that we targeted was lipid peroxidation. We had shown that lipid peroxidation in the injured cat spinal cord is a rapidly evolving process that begins within the first minutes after injury. We also showed that you could look at lipid peroxidation in terms of depletion of antioxidant levels, specifically vitamin E levels within the injured cord. Vitamin E is consumed as it quenches lipid peroxidation reactions, and by 4 h postinjury there is an 80% decrease in vitamin E in the injured cord, consistent with fulminant lipid peroxidation. We hypothesized that a steroid drug like methylprednisolone might be useful for inhibiting posttraumatic lipid peroxidation. In addition, we demonstrated that very high doses of methylprednisolone were needed, and that there was a peculiar U-shaped dose response curve. A lot of steroid was good, but a lot more was not better. We also picked a physiological parameter, posttraumatic decline in blood flow within the injured white matter of the cord, and we showed that the dose-response for the efficacy of methylprednisolone to reduce decline in the blood flow was the same dose-response curve that we had seen in inhibition of lipid peroxidation. We went on to show that blood flow decline was related to lipid peroxidation. When we pretreated animals with large oral doses of vitamin E, the expected decline in white matter blood flow was completely attenuated. We also looked at a marker of energy metabolism, tissue lactate, which is increased dramatically at 1 h postinjury. We showed that methylprednisolone could affect that increase, but a 30 mg/kg dose was required. Again, there was a peculiar U-shaped dose response curve. We showed that you could correlate tissue lactate levels in the injured cord with the levels of methylprednisolone within the injured cord. At the peak of methylprednisolone concentration after an i.v. bolus, there was a suppression of posttraumatic lactate accumulation. As methylprednisolone levels decreased, lactate rose. This result implied that a single dose was not adequate to take care of the pathophysiology, and that multiple doses needed to be given. We went on to show that indeed multiple dosing maintained the beneficial metabolic effect. We also showed this dose regime had a beneficial effect on neurofilament concentration. By 4 h postinjury, calpain-mediated degradation results in tremendous loss of neurofilaments. A single early dose of methylprednisolone had a small effect, but multiple doses that maintained the drug levels at an adequate concentration had an even better effect.

We went on from there to design a 48-h dosing regimen in collaboration with Doug Anderson and Gene Means, then at the University of Cincinnati. We showed that an antioxidant dosing regimen of methylprednisolone, facilitated neurological recovery in a cat model of spinal cord injury. Bracken et al. translated all this preclinical information into success in NASCIS II. The trial used high "antioxidant" dose levels of methylprednisolone for a 24-h period, and showed that, if treatment began within 8 h of injury, there was a beneficial effect.

We went on to develop Tirilazad, which is a nonglucocorticoid steroid inhibitor of lipid peroxidation. We did extensive dose-response studies in spinal cord injury models, and showed a benefit over a broad dose range. We showed that there was a 4-8 h therapeutic window, very reminiscent of the clinical trials with methylprednisolone. Subsequently, the NASCIS III trial looked at 24-h and 48-h treatment with methylprednisolone, as well as 48-h treatment with Tirilazad. The trial showed that all of these treatments had some effect and that 48 h of methylprednisolone was better than the 24-h regimen.

For TBI, we need to compare efficacy in multiple models, looking at both diffuse and focal injury, subarachnoid hemorrhage, and ischemia. The most thorough evaluation of a single agent for the treatment of head injury very likely needs to take place in all of these types of models. Another issue that needs consideration relates to possible differences in response for male and female animals. We have looked at this issue using the impact acceleration model of closed TBI. We find that during the first hour after injury (1.5-M weight drop, 500 g) there is about 25% mortality in males. There is no mortality in weight-matched females subjected to the same injury paradigm; however, there is a "male-like" acute mortality in ovariectomized females. We also examined blood flow in the injured cortex (laser Doppler flowmetry), and saw in males a massive, early drop in flow that was maintained over 90 min. Females showed a greater maintenance of blood flow in the cortex after injury, whereas ovariectomized female rats responded like males. Estrogen, the putative source of the post-traumatic difference, is a very complicated story; however, gender difference is one of the confounding factors that we need to take into consideration in future clinical trials.

I support the idea of trying to take the NASCIS II dosing regimen into human TBI. I still continue to believe that the antioxidants have great therapeutic potential. I also want to emphasize that we need to think about other approaches to the treatment of traumatic brain injury. Our focus has been on neuroprotection, but also we need to think about neurorestorative approaches. There are a number of potential strategies that we might use to try to induce structural plasticity within the injured brain. For instance, the compound OP1, which is a bone morphogenic protein, has been shown to facilitate recovery in stroke models with delayed administration until 24 h postinjury. Aiming at neural cell adhesion molecule (NCAM) antagonists or antagonists of the myelin-derived Nogo (ligand recognized by the myelin antibody, IN-1) protein are other approaches that I think need to be considered. Neurorestorative strategies might get us around some of the therapeutic window issues that have plagued acute neuroprotective strategies.

Question: Was the time window for methylprednisolone established in animal models of injury?

Answer: No. That was one thing that we did not do adequately. We did not do very complete therapeutic window studies with methylprednisolone in the cat model. We did have some data on delayed administration of methylprednisolone showing less of an effect after 2 h, but we did not look at it systematically.

Question: Do you think that the animal models represent the clinical situation?

Answer: Yes. I have long wanted to believe, as many of my colleagues here want to believe, that a 2-h therapeutic window in a rat perhaps translates to 8 h in the human. We do not know that. I think we have to take the animal therapeutic window seriously as we plan clinical trials, because we do not really know that they are different.

Preclinical Testing of Drugs (Tracy McIntosh, Ph.D.)

The speakers this morning did not blame our current disappointment with clinical trials on a lack of fidelity of the animal models to clinical TBI. Head injury is difficult to model because of its heterogeneity. It is admittedly very challenging to develop a rodent model, or other animal model, that possesses all of the features of human injury. Experimentalists acknowledge and respect the difficulties regarding clinical trials: the heterogeneity of patients, and the challenge of demonstrating efficacy with a single compound, using a single dose. We face many of the same problems in the laboratory. What then should be considered when modeling TBI? One issue has been the relative lack of funding for vigorous preclinical trials. Many sponsors are reluctant to support preclinical trials that address the relevant mechanisms, such blood flow,

metabolism, and receptor changes. The emphasis has focused solely on efficacy with one dosing regimen, one drug, and one outcome measure. This narrow approach can be frustrating for investigators who have spent years documenting the pathobiology and clinical relevance of these injury models. We need to recognize that there are several reasonably good preclinical models of TBI; then, we need to agree on which ones to use, use them collectively, and share data.

Most laboratories are all working with models that include both focal and diffuse injury. However, we know that the pathobiology of a focal injury is different from that of a diffuse injury. There is no model of inertial injury (inducing diffuse axonal injury without focal contusions) that is available to all laboratories; but there are excellent, relevant models of mixed diffuse and focal injury. A number of these models can be used successfully to mimic those mixed injuries that one sees clinically. Decades ago the existing experimental TBI models were not clinically relevant, so people relied on preclinical data from other diseases. But now several of the models have been characterized and are believed to accurately mimic the heterogeneity of human TBI. It would be wonderful if we could all agree to use these pre-clinical head injury models as the basis of our clinical trials.

What preclinical criteria should be standard or accepted before we move on to clinical trials? Clinical trials have at times gone forward based on results from one laboratory, or without any preclinical data from a relevant TBI model. We must use multiple models (focal, diffuse, mixed) to evaluate these compounds before they go into patients. Moreover, in the laboratory we most often use models of moderate, or mild/moderate head injury. Even if clinical trials are selecting drugs that we have tested preclinically, but go forward to test them in severely head injured patients, are we making a mistake? Should we widen the scope of trials and test the preclinically effective drugs in mild/moderately head-injured patients?

Another topic that Ed Hall emphasized is that of "polypharmacy." Currently, there are few studies performed in TBI laboratories that combinations of drugs. It is very difficult to test the interactions of multiple drugs in the laboratory, but we must recognize that we have seen the list of pathogenic factors that are active in TBI increase over the past decade, and continue to get longer. One treatment is not likely to inhibit or attenuate all the many pathological mechanisms. Dose-response and therapeutic critical window studies must be performed. I suggest that clinical trials not begin unless the compound or proposed treatment has been tested in one or more valid preclinical models of head injury, by several reputable laboratories looking these issues.

Question: There are many studies that showed protection utilizing so-called valid preclinical models that have not translated to the human setting. How can we respond to that?

Answer: Two possible reasons are the differences between severities of injury and the means of outcome assessment. Many preclinical studies that have been used as the basis for clinical trials have *not* used behavioral endpoints as their outcome measures. As we know, clinical trial endpoints are predominantly behavioral. There needs to be more consistency between the pre-clinical findings and the clinical studies. If you want to study the effect of a compound on ICP/edema in head injury patients, then positive ICP/edema data should be generated first in some relevant preclinical model. If you want to assess a GOS at 1 year, it would be inadvisable to take a drug to trial that has only been shown to reduce acute edema. Long-term neurobehavioral endpoints are extremely important in the preclinical data.

Question: What is the best animal model?

Answer: The lateral fluid percussion model in the rat has been the most successful and widely used preclinical model to study acute pharmacological intervention. The pig brain is more

gyrencephalic and similar to humans, but no one has developed suitable behavior tests for functional evaluation in this species. We are currently limited to the rodent models with respect to behavioral endpoints.

Outcome Measures (Graham M. Teasdale, M.D.)

There is not a gold standard, numerical outcome measure for "behavioral outcome." It is always necessary to compare against something, and different perspectives can account for the different outcomes that people want. Perhaps we want to compare people with their healthy selves: are patients back to what they were before injury? That is difficult to assess in an individual. We tend to compare against population norms, but head injuries are not a random samples of the population. Those who treat these patients know how bad they were in the early stage, so that patient survival is a kind of triumph. Yet that clearly is not sufficient. What is the natural history of TBI: are the patients getting better faster than without treatment?

For any particular clinical question, the outcome measure must be relevant to the question, must be robust enough to be used widely, and must influence people to decide that your findings are valid. The GOS is still the most widely used index, more than 25 years after Jennett and Bond described it. It has withstood the test of time because it has many useful attributes. One of the things that have happened over the years is that people have tended to misuse it, and think of it as a scale for physical outcome. It is not. The authors emphasized social aspects of outcome. Really, it is more of a handicap scale than a disability scale, determining a person's independence in society and return to previous lifestyle.

Many years ago, the categories of survival that makeup the GOS were subdivided into upper and lower degrees of severe and moderate disability and good recovery, but the approach did not really catch on. Consequently, the five-category scale is the most widely used outcome scale, and has been for two decades. Whatever we do we must keep the GOS in the picture, or we cannot compare what we do in the future against history.

To some extent, the attributes of this scale can be disadvantages. It is so easy to use that it can be misused, and it can be done so quickly that it can be done superficially. It is open to varying interpretations; and doctors tend to underestimate disability, compared to what a rigorous discussion with the patient or caregiver by a psychologist gives. Professor Lindsay Wilson and I have described a structured interview for allocating people to the different GOS levels, getting away from the subjective approach that had been misused over the years. The approach sets out the questions that should be asked, and the answers provide a guide to the allocation of subjects on the original five-point scale or on the extended eight-point scale of the GOS. The structured approach dramatically increases interobserver reliability. Both original and extended scales show very strong correlations with a host of neuropsychological measures. More recently, we have developed a postal version of the questionnaire that shows very good correlation with a face-to-face interview.

Another issue concerning outcome is how outcomes distribute in real life. Is the "U-shaped" distribution real in a series of head injury patients? A large proportion die and, of those who are alive at 6 months, most have made a good recovery. With the customary approach to dichotomization (D/V/SD vs. MD/G recovery), the division is at a point where there are very few patients. It is most difficult to show a treatment effect.

Maas provides an excellent paper that discusses the use of GOS in TBI trials. Various suggestions are made on dichotomization of the scale, statistics, and patient selection. Narrowing selection to exclude the most severe as well as the mildest injuries has the advantage of obtaining the same power from a reduced number of recruited patients. This is statistically more efficient and may also be biologically more efficient.

Nevertheless, I wonder about the message that the benefit of a new treatment is that it increases the number of people with moderate disability. That kind of result is relevant, but how influential and how generalizable will it be? The approach to dichotomization needs to be examined, and the dichotomization point set differently for different severities of initial injury. Splitting severe versus moderate and better, the classical unfavorable versus favorable dichotomy is probably relevant to severe head injury, as we have always defined it. In the milder head injuries, a better discrimination may be moderate disability versus good recovery at 1 year. In an unselected series of mild injuries in Glasgow, we found that at 1 year only 50% of subjects make a good recovery. By shifting the dichotomization point from severe/moderate to moderate/good, a valuable effect in mild injury can be shown with the GOS.

To test this approach, Murray took into account age, coma score, pupils and CT scan findings to produce four broad "prognostic bands." The approach was applied to the patient population in the EBIC core databank. This produces a clear relationship between the outcome distribution and the early prognostic band. Then, in each band, the population was dichotomized into approximately equal halves. In the poorest prognosis band, the relevant dichotomization is between upper and lower severe grades on extended GOS (GOSE). For those with not quite as bad a prognosis, the dichotomization for outcome might be drawn between upper and lower moderate disability; for those with better prognosis, dichotomization falls between moderate disability and good recovery. With a population at entry of mostly mild TBI patients, dichotomization could move to between upper and lower good recovery. This may be a way of using the data to maximum effect, building in early prognosis, and relating that to the expected effect of treatment. This "roving dichotomy" gets away from the traditional split in a way that clinicians will find intelligible and relevant.

Early indices of effect (so-called surrogate outcomes) may be very interesting biologically, may be very sensitive, but may not predict the eventual clinical benefits. For dose-ranging or proof-of-concept studies, surrogate indicators might be useful. If the interest is general clinical adoption, there must be some index of outcome related to quality of life and function. It would be no surprise that I am skeptical about the value of ICP measurements as a surrogate indicator. There are now studies that show discrepancies between effects on ICP in the early stage, and ultimate patient outcome. In a classic study on hyperventilation, high ICP was more common in the control group, but outcome was actually better. Two similar examples have been presented at this meeting. In Robertson's study, the ICP profiles were better in the group with worse outcome. In Clifton's trial, there was a 20% reduction in ICP.> 30 mm Hg in the hypothermia group, but no difference in outcome. As a surrogate index of outcome, ICP clearly has flaws.

Question: Could you address the role of imaging in assessing outcome?

Answer: If the question concerns the degree of swelling, then magnetic resonance (MR) or CT would be fine. If your question relates to overall functional outcome there are problems. In head trauma there is no visible MR index that correlates with the degree of axonal injury. If you are going to use MR to look for changes, I think you must use magnetization transfer or another sort of numerical index. You need a measure of white matter injury that will respond to an intervention and that any change will correlate with later outcome. I think people have misused the term, "surrogate outcome measures"; they want to use it as the sole indicator of whether treatment is good or not. I think it is better used as a way to determine if a treatment is effectively targeted.

Question: I am troubled by the experience in AIDS research, where it took 10-12 years of unfocused work, lacking progress toward treatment, until researchers managed to crack the issue of finding the best surrogate indicators of their disease, the CD4 counts of viral load.

They had a number of false leads before they could define a measure of whether a therapy is going to make a difference. We have to keep looking for that particular surrogate indicator for TBI.

Answer: In fact, I think that there are many early markers that correlate with outcome, but we have to go beyond that conceptually. The desired early marker must correlate with the effect of treatment on outcome. Such a correlation remains to be demonstrated.

Selecting Outcome Measures (Nancy Temkin, Ph.D.)

In order to have a successful trial, patients need to have room to improve. Many studies target severe injury based purely on GCS and select a primary outcome of good recovery/moderate disability on the dichotomized GOS. For severe TBI, about 64% of patients in our studies do well at a year, versus close to 100% in trauma (not nervous system) controls; that leaves only 36% of patients who could improve. If you include moderate injuries (GCS 9-12) there is only about 13% room to improve on a dichotomized GOS as the primary endpoint. If you use other endpoints (for example, performance IQ or return to work) you can use less severe cases, and still have plenty of room for these people to show improvement.

What sorts of outcome measures are appropriate? Using a simulation for the population of the magnesium sulfate study, we plan for 400 cases; both moderate and severe head injuries. Inclusion is not limited to GCS 8 or less, but goes up to GCS 12 or less and anyone who has an emergency craniotomy. For our simulation, the difference to be detected is 10% more subjects in the "good" range on a dichotomized GOS. For the severe cases, there is 36% room to improve on that outcome; 13% room to improve for patients with moderate injury. For the dichotomized GOS in this trial, there is one chance in four of having a successful trial to show the difference. If you use the full GOS, all five categories, power increases to about 50%, one chance in two of having your trial have a significant result. If you choose performance IQ, power goes to a 66% chance of seeing a difference; for selective reminding recall (a test of verbal memory), the power is 99%. Based on these calculations, our primary endpoint is a composite measure of survival, seizures, functional status measures, and neuropsychological measures. We have chosen a different approach to some of the same problems brought up by others by targeting a broad population with a more sensitive outcome.

Question: What are priorities for research? How much need is there for identifying good measures of patient assessment verses the need for better therapeutic agents?

Answer: There is a need for both. There are two aspects to any assessment question. There is assessment of severity and type of injury in order to target the population and to make sure your treatment groups are balanced, either by design or by modeling in the analysis. There is also the need to assess outcome, and I think that the measures exist. We have a lot of data on assessment tools that have not been used in clinical trials. I think there are unfortunate perceptions, many unfounded: that the FDA will not accept them, that they are too difficult, that you cannot train people to give them, patients cannot take them. Absolutely, it does take significant oversight to obtain good neuropsychological data. You cannot provide a one-page instruction sheet. Examiners must learn the measures, and know them well; that is a problem if you have two or three patients per year at a site. But, if you can do trials where you have even a moderate number of patients at a site, you can use much more sophisticated outcomes than the dichotomized GOS. Neuropsychological testing seems to be the most powerful but there is also the GOSE, which has recently been published. We have just submitted a paper on functional status examination. The note of caution for any outcome is that you must take into account in the power calculation that you will not have 100% follow-up even under ideal circumstances.

Question: I found your comments regarding the "neuropsych" measure very interesting. How do you assess patients who are severely injured, with a much less favorable outcome, where you cannot test?

Answer: For patients who die, we assign the worse possible score. Some patients, badly impaired or vegetative, cannot do the tests. We assign them one better than the least score, but worse than anything that can actually be assessed. You must specify all these things in the protocol, but it is not an insurmountable problem.

Question: I think the return to work measure is very important. Many TBI survivors are more incapacitated by their behavioral difficulties than by their cognitive difficulties. Are you looking also at sustained work, that is, staying at work more than 180 days? What do you do about the substantial portion of patients who are not gainfully employed at the time of their injury?

Answer: We did not look at retention, just whether patients were working or not at the time of their 1-year assessment. We collect information on the date that they went back to work and their entire work history until the end of assessment, 2 years for most patients in our trials. In our experience, about 20% of the control and treatment cases were not working.

Choice of Primary Endpoints (Sung C. Choi, Ph.D.)

A biostatistician's view of appropriate primary endpoints might be different from a neurosurgeon's. Prerequisites for an ideal primary outcome measure include the following:

1. The measure should be ordinal. "Ordinal" means that the measure must have inherent rank order in its scale.

2. The measure should be able to be interpreted in the magnitude of a potential treatment effect.

3. The misclassification rate should be minimal.

The most widely used outcome measure in clinical trials of severe head injury is the GOS. The scale consists of five categories: good recovery (GR), moderate disability (MD), severe disability (SD), vegetative (V), and death (D). The GOS is ordinal. In most trials, the GOS is dichotomized to favorable (good) outcome by combining GR and MD, and combining the other three categories as unfavorable (poor) outcome. Making SD as a separate third category could trichotomize the GOS. The GOS also could be used as a four-category outcome measure by combining only V and D.

Another outcome measure, advocated by some investigators is the DRS. The DRS ranges from zero to 30, where zero is the best score and 30 is death. The measure is the sum of eight individual items. Clearly, each individual item is ordinal but the DRS would be no longer strictly ordinal. For example, a patient could have a better DRS than the second patient simply because he/she has much better score in communication while the second has a better score in feeding. Even so, the DRS is approximately ordinal (i.e., stochastically ordinal) in the sense that if one patient has a better score than a second patient, the overall outcome of the first patient is likely to be better.

The Functional Independence Measure (FIM) is another well-known, potential outcome measure. This measure might be considered an extension of the DRS. The FIM is a sum of 18 items, with each item scored from 0 to 8. Like the DRS, the FIM is not strictly ordinal. A measure such as the DRS and the FIM, which is made up many items, can be called a composite measure. In general, composite measures can be only approximately ordinal, and therefore, the measures do not possess the property of a desirable measure.

The primary advantage of the GOS is that it is both easy to interpret and can quantify the potential efficacy of a treatment. The interpretation can be difficult with the DRS or the FIM. Suppose that the treatment group has a better average score than the control group by seven points, in the mean or median of the DRS. The seven points could indicate various degrees of difference in outcome. For example, the seven points could be the difference between two vegetative states or that between no disability and a moderately severe disability.

Next, we must consider the effect of misclassification. The following is an example of hypothetical outcomes of 800 patients (Table 2).

Suppose every patient has a 10% chance of being randomly misclassified to adjacent categories. Among 240 patients in the treatment group with good outcome, 24 would be misclassified to poor outcome, while 16 with the poor outcome would be misclassified as good outcome. Thus, the observed number of favorable outcome would be 232 instead of 240. Similarly, the observed number of good outcome in the control would be 208, down from 210. As shown in Table 2, the difference in outcome between the two groups decreases and the power is reduced. The power decreases as the misclassification rate increases. It has been theoretically shown that the loss of the power due to misclassification can be substantial.

Clearly, the misclassification rate of an outcome measure with a large number of categories is likely to be higher than in a measure with a smaller number of categories. For example, it is likely that misclassification using the three category GOS would be higher than in the dichotomous GOS. Moreover, the misclassification rate of the DRS or the FIM is expected to be higher than in the GOS. It should be noted that a part of the problem with the DRS and the FIM is that each one of the many component items can be misclassified, and that could have cumulative effect on misclassification.

Finally, we consider the results of analysis of a subgroup of patients from the recently conducted hypothermia trial. Briefly, 392 patients aged 16 to 65 with severe closed head injury were randomized to hypothermia (33°C for 48 h) or normothermia. The subgroup consisted of 81 patients with age ≤45 years and the admission temperature of ≤35°C. The significant probabilities (*p* values) based on the Wilcoxan rank-sum test using four different outcome measures are presented in Table 3.

We expected that the DRS would yield the most power, and the dichotomous GOS would be the least powerful. The results are completely the opposite. For this particular subgroup, the dichotomous GOS is the most sensitive outcome measure. The reason for the seemingly anomalous results becomes clear if the distribution of the outcomes is examined (Fig. 1).

Note that the effect of hypothermia is mainly confined to MD and SD, while the three other categories (i.e., GR, V, and D) are not significantly affected by the treatment. In the subgroup, the power decreased because all other categories became the noise in the analysis. What this example shows is that the power does not necessarily increase as the number of categories is increased. Another partial cause of the decreased power with more categories in this example may be the increasing rate of misclassification, and in addition, nonordinality in case of the DRS.

In conclusion, well-known composite measures such as the DRS and FIM may be not ideal as the primary outcome measure for clinical trials of severe head injury. Such measures are often used as the secondary outcome measures. These measures were shown to be effective in monitoring the same group of patients over time. The measures may also be appropriate in trials of moderately injured patients in whom the GOS is simply not sensitive enough. It is important to recognize that a larger number of categories do not always yield greater power. The ideal number of categories is likely to depend on the effect of treatment, the

misclassification rate and the shape of outcome distribution. If the outcome distribution is "U" shaped, as in the case of severe head injury, a smaller number of categories might be better than a large number especially when the misclassification is taken into account.

Question: Scales like the DRS and FIM have not been designed to create multi-dimensions. But the scales supposedly have an ordinal and even an integral structure.

Answer: Some investigators think the DRS and FIM are ordinal because there is a numerical order in the measure. Numerical order alone does not make a scale ordinal. As stated, the DRS and FIM are approximately ordinal.

Question: Some of the available clinical data would show patterns of effects; however, the dichotomous GOS would not show any difference.

Answer: If all categories of the GOS are affected by the treatment, then the test based on the dichotomous GOS would be less powerful than using four or five categories. The dichotomous GOS would yield more power when the primary effect of treatment is more or less localized to two categories, as in the example of the subgroup in the hypothermia trial.

Question: I think your point about misclassification is exactly right. Misclassification is always a problem, and if it is random it will reduce the power of the study. I am a little concerned at your assertion that DRS and FIM are not ordinal. I think that if that criterion were applied, the IQ would not be ordinal. The predicted value from a logistic regression with more than one variable would not be ordinal, because it is made up of a combination of parameters.

Answer: Some composite scores may be more ordinal than others. I believe the DRS is less ordinal than the IQ because IQ is composed of fewer items than the DRS. Similarly, the FIM is likely to be less ordinal than the DRS. The predicted probability from a logistic regression is most likely ordinal because the regression coefficients and the prediction are mathematically determined from the data, such that the predicted values based on the regression are order preserving. The other point with the composite score is that it is not clear how different items are weighted by giving different range. The weight assigned to each item seems subjective.

Question: You indicated that the number of misclassifications is linearly proportional to the number of categories. What is the evidence for that?

Answer: Clearly, if we create more categories by dividing some or all of the categories, misclassifications will tend to increase. In the GOS, for example, if we simply dichotomize, we would not have misclassifications between good recovery and moderate disability.

Realistic Expectations from a Traumatic Brain Injury Drug (Anat Biegon, Ph.D.)

Neuroprotective agents or therapies are defined by their ability to prevent neuronal death in groups of neurons damaged distally at the time of the original insult. This is achieved through inhibition of physiological processes (e.g., excitatory neurotransmission, cytokine release, free radical formation, and temperature regulation) that are acutely overstimulated following brain injury. Thus, a neuroprotective agent is expected to improve outcome by preventing the propagation of injury to remote brain regions, in contradistinction to neurorestorative interventions, such as growth factors and transplants, which are expected to improve outcome by replacing dead neurons. Based on this theoretical framework and on what was considered to be relevant and encouraging pre-clinical evidence from animal models, several neuroprotective agents and procedures (e.g., Tirilazad, Selfotel, Cerestat, hypothermia, D-CPP-ene) were tested in phase III clinical trials in TBI and stroke. All of the above trials failed to show efficacy of the test drugs or procedures, and in some cases the outcome of the treated

group was actually worse than the outcome in the control group (e.g., hypothermia in TBI, Selfotel in stroke). The following detailed examination of the gap between the expectations from neuroprotection based on animal models and the results in human clinical trials may contribute to our understanding of the failures in the clinic, a well as to the design of more appropriate preclinical testing and better clinical trials in the future. Table 4 summarizes the features of the latest neuroprotection trials and the preclinical data used at the time to support their testing in humans.

It is easy to see that the clinical trials and animal studies differed significantly in design and execution. In general, the conditions under which the agents were tested in humans have never been validated in animal models. Minimizing the gap using the strategies outlined below may increase the likelihood of positive clinical trials in the future.

The lack of efficacy in humans can be largely attributed to the unrealistic therapeutic window (time between injury and effective initiation of treatment). Animal models invariably show that glutamate antagonists, antioxidants, and hypothermia are most efficacious when administered before or during the insult. The efficacy of these interventions drops precipitously with time, with no effect seen for treatment delayed by more than 1-2 h. To illustrate, Cerestat was never tested later than 15 min post insult in a TBI model, and the hypothermia literature in animal models showed conclusively that delayed application of hypothermia is not neuroprotective. While it is theoretically possible that the treatment window in humans might be somewhat longer than in animals, there is no data currently available to support this notion or to otherwise define the relevant treatment window in humans. The clinical experience with tPA, demonstrating the drug is not efficacious in human stroke when given more than three hours after onset, actually suggests that the treatment windows for thrombolysis in animal and human stroke are quite close. An optimistic expectation that the time windows in human TBI will be considerably longer was not supported by the results.

While late intervention can explain lack of efficacy, it cannot explain unexpected toxicity. The detrimental outcome of neuroprotective treatment observed in some of the clinical trials (Selfotel/stroke, Cerestat/stroke; hypothermia/TBI) is possibly attributable to the difference in treatment duration. Blockade of NMDA receptors, cytokine release, or core temperature regulation beyond the therapeutic window appears to be useless and can be risky since it might interfere with recovery. A growing body of literature suggests that normal levels of NMDA receptor stimulation and TNF-alpha production are essential for neurodevelopment, CNS tissue remodeling and plasticity. These processes are the basis of spontaneous recovery of function.

In support of this explanation, in a recent experiment we have shown that repeated (6 days) administration of the potent NMDA antagonist MK801 in animals with middle cerebral artery occlusion resulted in a significant increase in the number of deaths over a 60-day follow-up period, when compared to a single administration. Mortality rate was 23% in vehicle-treated animals, 0% in the single treatment group, and 38% in the repeated treatment group.

Similarly, Work from the McIntosh laboratory has demonstrated that TNF-alpha knockouts have better short term outcome (48 h), but worse long-term outcome in a brain injury model compared to wild-type animals. TNF-alpha knockouts actually simulate a chronic blockade of the cytokine; and provide another example for the risk involved in long-term blockade of processes that are only activated acutely following brain injury. It is important to note that the hazards of overlong treatment with neuroprotective agents cannot be predicted from repeated administration paradigms in toxicological studies performed on healthy animals or volunteers, since the risk may very well derive from interfering with recovery processes that only occur in the injured brain.

A common outcome measure in animal models of brain injury is infarct size at 1-3 days postinsult. Recent human and animal data suggest this is not a valid or predictive endpoint. Results from the RANTTAS (Randomized Trial of Tirilazad in Acute Stroke Patients) stroke trial show a poor, nonsignificant correlation between infarct size on CT in the early days after stroke and clinical outcome after three months. Short- and long-term outcome studies in animal stroke or TBI models are rare, but data from spinal cord injury models in animals had shown instances of dissociation between short-term reductions in infarct size and long-term functional outcome. These observations underscore the need to include functional ("clinical") endpoints in future pre-clinical studies of neuroprotective agents.

The clinical outcome measures commonly used in clinical trials in TBI are not problem-free either. The primary outcome measure for the majority of recent clinical trials in TBI required a 10% shift towards favorable outcome on the dichotomized GOS. When this requirement was formulated, severe head trauma (described in the seminal paper from the traumatic coma data bank) entailed a mortality of more than 40% while good neurological recovery was relatively rare. Over the last couple of decades, a decrease in mortality and shift towards good outcome has been observed in all recent clinical trials, such that a 10% difference in favorable outcome is mathematically impossible to achieve. Thus, once the mortality rate drops below 20% in placebo treated patients (e.g., the Selfotel trial), a 10% shift in favorable outcome can only be achieved if the drug reduces mortality by more than 50%. This problem can be overcome if the patient population is heavily weighted towards poorer outcome by imposing inclusion criteria such as GCS, 7, CT. 2, or presence of SAH, but this may slow enrollment considerably. On the other hand, detection of a 10% or more improvement in good recovery is quite possible in the broader severe head trauma population (GCS 4-8, CT 2-6, with or without SAH) as well as in the subgroup of the less severe patients (GCS 7-8, CT 2).

Finally, assessment of clinical improvement by global scales such as GOS assigns unequal weight to rescuing neuronal populations of equal size, thus contributing to the extreme variability in outcome. Thus, damage to the optic nerve will result in blindness (bad outcome on GOS). Severing of the entire corpus callosum, a much bigger axon bundle, will not be detected on GOS at all. To decrease outcome variability, a scale based on regional/functional systems (similar to EDSS [Expanded Disability Status Scale]) is more likely to reflect the expected effect of a neuroprotective agent, that is, rescue of neuronal populations at risk anywhere in the brain.

Data gathered from a variety of animal models of TBI and stroke demonstrate that neuroprotective agents administered within their therapeutic window have a reliable and highly reproducible effect on early outcome measures such as infarct size or neurological function at 24-72 h postinjury. These results cannot be taken to imply persistent long-term (3-6 months) benefit in animals or humans. The only way to predict the long-term effect of neuroprotective agents is to increase the length of follow-up in the animal studies such that it is comparable to the intended follow-up period in clinical trials. Interestingly, in the very few studies evaluating the effects of neuroprotective agents on functional, long-term outcome available to date, it appears that the duration of the benefit depends on the severity of the initial insult. Thus, in moderate or mild insults, the early treatment benefit diminishes and disappears as control animals recover spontaneously over time. A stable, long-term benefit of treatment is only observed in more severe injury models, such that full recovery is not feasible with time alone.

Taken together, these observations suggest that neuroprotective agents can be relied on to accelerate recovery in all patients, but improve final outcome only in those least likely to achieve good recovery without additional intervention beyond standard care, rehabilitation, physical therapy etc. Thus one could expect neuroprotective agents given to humans to reduce neuronal cell death and improve neurological function at early time points, such that treated

patients will regain consciousness, leave the ICU, leave the hospital and reach their final level of improvement faster than untreated patients. In terms of human suffering and medical costs, acceleration of recovery might be almost as important as a difference in long-term outcome. Therefore, it may be beneficial to add early outcome measures to future clinical studies.

Obviously, patients within clinical trials should not be denied any supportive or concomitant treatment that is part of the standard of care in TBI. For a preclinical study to be predictive of efficacy in the clinic, the drug/intervention need to show efficacy in animal models in conjunction with essentially similar supportive treatment, such as maintenance of blood pressure, resuscitation etc.

Preclinical studies need to include both sexes and young as well as old animals to be predictive of outcome in the broad patient population.

A critical review of the available pre-clinical and clinical data recently obtained with neuroprotective agents leads to the following conclusions and recommendations:

- Neuroprotective agents should be tested within the therapeutic window associated with their mechanism of action in animals or, if the data are available, in humans.

- Treatment duration in clinical trials should not exceed that which was found to be effective and safe in animal models of TBI or stroke (not healthy animals or volunteers).

- Preclinical studies should include long-term functional outcome measures and encompass more genetic variability, both sexes and a variety of age groups.

- Clinical trials of neuroprotective agents should emphasize speed of recovery by including appropriate measures of this parameter as primary outcome criteria.

- Dichotomization of GOS between "good recovery" and "not good" (MD, SD, V, D) better reflects recent trends in outcome and has the potential to produce larger group differences and smaller sample sizes.

- A new scale for measuring brain system improvement, constructed from EDSS with the addition of items for cortical and subcortical regions gleaned from recent functional imaging studies, employed repeatedly over time, might offer an advantage for assessing the efficacy of neuroprotective agents.

Alternative Trial Designs (Charles Contant, Ph.D.)

We are generally familiar with the nomenclature of clinical trials as phase I, II, III, and perhaps IV. Piantadosi has proposed an alternative naming system that I feel is somewhat more descriptive. We seem to have the perception that, in our field, we start with six mice and then do a trial in 600 people. While this is hyperbole, I would like to discuss the types of studies that occur between six and 600. These are the phase II trials, or the safety and efficacy studies discussed by Piantadosi. Phase II studies are performed to determine if evidence exists to continue the development of the treatment. These trials may be performed to determine the safety profile of the treatment, to determine the safest and most effective dose (dose finding or dose ranging), and to evaluate whether the treatment is efficacious enough to warrant further study. These types of studies should be performed and evaluated before the commitment of resources to a large randomized clinical trial. I do not believe that patients should be exposed to a treatment unless there is evidence that the treatment works in humans at the dose selected without causing harmful side effects.

There are a large number of perfectly acceptable phase II designs. I am often asked if a recipe exists for phase II trials. Unfortunately, no template is available; most phase II designs are

crafted to answer the question being asked. There are broad classes of designs. One important caveat for the designs that follow is that most require that the trial response variable can be quickly determined. Few of the designs are appropriate for the 6-month outcome assessment time used in most phase III TBI trials. Therefore, one is usually assessing a surrogate indicator of the 6-month outcome of interest. Such surrogates require much further investigation. One might argue, however, that if the treatment does not have an effect on a pathophysiology induced by the injury, it has little likelihood of affecting the long-term outcome.

Fixed sample size trial. These studies are performed to estimate a property of the treatment to a certain precision, or to determine if the treatment is better, in some sense, than a "historical" control. In the first case, an example would be to estimate the rate of a serious adverse event so that the 95% confidence limit around the rate was only 10% (65%). The actual rate is of less concern than the precision with which it is estimated. The second circumstance might occur as follows. Assume that the current consensus is that a treatment works in approximately 40% of the patients, and that the new treatment would be useful if it works in 60% of the patients. Under these conditions, only 38 patients are needed to determine if the rate is about 40% at the alpha level of 0.05. This evidence is not sufficient to conclude that the drug is actually better than the historical record, but only that the drug is probably worth further development. The fixed sample size designs can be appropriately used for long-term outcomes.

Staged designs. The most versatile type of phase II trial is the staged design. In this type of trial, one or more decision point is designed into the study at which the study can be continued or terminated. For example, the trial may be conducted until a certain number of patients have received the new treatment. At that time, the outcomes are evaluated. If the treatment has performed up to a prespecified design criterion, the study continues in a second group of patients until the final sample size is reached. If the treatment does not reach the design criterion, the study is terminated and the new treatment is rejected as not worth further consideration. The most well known staged phase II design is that by Simon (Simon, 1989). This two-stage design is based on a binary response (e.g., success/failure). The historical and new treatment response rates are prespecified. Assume that these rates are 0.40 and 0.60 as above. Unlike the fixed sample size example, these rates define three regions. If the treatment success rate is 40% or lower, the drug is not worth further development. If the rate is 60% or higher, continued development is assured. Between 40% and 60% is the "indifference zone." Should the treatment success rate fall in this zone, factors other than treatment efficacy will be used to determine further development. Staged designs require fairly rapid assessment of outcome.

For a Simon two-stage trial comparing 0.40 to 0.60, the study proceeds as follows. Initially, 16 patients are enrolled and evaluated. If seven or fewer patients have successful outcomes, the trial is stopped and the treatment rejected. If eight or more respond positively, a second group of 30 patients are enrolled and evaluated. From the final total of 46 patients, if 23 or fewer have successful outcomes, the drug is rejected. Otherwise, the drug is accepted for further study. Under this type of design, the sample size is either 16 or 46, with the average being just 25.6.

Staged trials can also be designed for evaluating complications, for selecting among doses, or selecting among treatments. A staged-hybrid of the phase II and III trial exists where the information from the phase II design is used to determine the sample size of the phase III trial, and all the patients are included in the final analyses. There is clearly a wealth of possible staged designs, which I believe should be used in the treatment development process.

Selection trials. Selection designs are for a very focused group of phase II trials that are designed to select a group of the best "treatments" from a larger group of possible treatments.

The best known of this type of design is the "dose-finding" study, in which the "treatment" is the dose of the drug to be delivered. One potentially useful definition of "treatment" is clinically distinguishable subgroups of the population. As we have head in earlier talks, focusing a phase III trial on those patients who have the greatest likelihood of responding to the treatment is one way of improving the efficiency of a trial. Designing a study to select these groups from the larger population may be worth considering. For example, a trial to select the best subgroup from a group of four subgroups so that there is a 95% chance the actual best subgroup was selected and that the response rate in this group is above 20% requires 53 patients per group. While this seems like a large trial (212 total), if the gain in power in the final trial is large, the total sample size for the two studies may actually be lower than if the selection had not been performed. Selection trials can be used with long-term outcomes.

Sequential clinical trials. Sequential clinical trials fall into two distinct groups: grouped sequential and fully sequential. Grouped sequential trials examine the data at fixed proportions of the total accrual, or at fixed times. Interim analyses are examples of grouped sequential designs. Whitehead has described the fully sequential designs extensively. Technically these designs are not phase II designs, as they require a randomized control group. They also require an outcome that is rapidly available, and are therefore only suited to the "surrogates" of the long-term outcome. I am presenting them here because they can be conducted fairly quickly and with potentially much smaller sample sizes than a full-scale phase III trial of long-term outcome.

Whitehead also proposed triangular designs. These designs are based on selection of a region on a graph where the null hypothesis is rejected, one where it is highly likely that the null hypothesis will not be rejected, and a region where the current information is not sufficient to decide. Consider a study that has as its outcome a simple binary assessment of whether the patient experienced transtentorial herniation during the first week following severe brain injury. The horizontal axis is proportional to the amount of information accrued, and expressed as the total sample size. The vertical axis is the difference between the two groups. The area within the triangle defines the region where there is not enough information to make a decision. The rest of the graph defines areas where a decision has been reached. The procedure is to graph the statistic defining the current difference between the two groups against the current information, or sample size. Initially, the trial is in the "undecided" area. As the trial progresses, the graph is updated with the new information. If a new point lands outside the "undecided" region, a decision point has been reached. This is called crossing the boundary. If the boundary is crossed between points A and B, the decision is reached that treatment is better than the comparison. If the boundary is crossed between points C and D, the decision is reached that the comparison group is superior to the treatment group.

The sample size required for the herniation study, using a classical fixed sample size phase III design, is 194. The sample size of the triangular design is actually a random variable. The average sample size is 100 at the null hypothesis, and 128 at the alternate. Fifty percent of the trials run using this design will require 92 or fewer patients if the null hypothesis is true, and 123 if the alternative is true. At the null hypothesis, 90% of the trials will terminate with 160 patients; at the alternate hypothesis, 90% of trials will terminate with 199 patients. Clearly, the sequential design has a high likelihood of terminating sooner than the fixed sample size trial. The worst case (rare, but possible) is that this design might actually require far more patients, up to 296.

Sequential designs best illustrate why the outcome should, preferably, be rapidly available. At any point in time, there are patients who have been randomized but have not reached the end of the assessment period. If two patients are randomized a week, a 1-week outcome assessment period leaves only two patients in "limbo." If the assessment period is 6 months, this number

increases to 48 patients. This large number would have a large likelihood of crossing back into the undecided region at some point in the six months after stopping the trial. While it is possible to include this scenario in the original design of the trial, it results in a greatly inflated sample size requirement.

Summary. I would argue that we ought to be doing more phase II studies in the development of treatments for TBI. Before proceeding to a phase III trial requiring 800 patients, we perhaps should have better evidence that the treatment shows some evidence of efficacy. A phase II design can be performed with relatively few patients, and might help avoid the need for large, and expensive, phase III studies using treatments that do not work. While the designs can be complex, collaboration between clinicians, clinical trial specialists, and biostatisticians can produce efficient and practical designs. I would argue that more of this collaboration is needed.

Question: Given the problem of heterogeneity in TBI, there might be patient characteristics that are not independent events.

Answer: If you mix a population of people in whom the drug will work with a population where it will not, you need a much larger sample size to see the effect. A selection trial may be a way to focus on the group that is most likely to benefit from the treatment.

A Call for Larger, Simpler Trials (Michael Bracken, Ph.D.)

Over 200 randomized trials of therapies for TBI are in the literature but no therapy has been shown to be effective. The Cochrane Library (1999, issue 4) includes six systematic reviews of some of this body of evidence— for barbiturates (five trials), calcium channel blockers (four trials), corticosteroids (19 trials), hyperventilation (one trial), mannitol (three trials), and therapeutic hypothermia (eight trials). All conclude that the therapies studied were either ineffective or that evidence of efficacy remains inconclusive. This presentation focuses on methodological issues in phase III trials of pharmacological agents since these have been most commonly conducted.

It is generally agreed that there is unlikely to be a "silver bullet" for head injury, that is, a therapy that influences a broad range of head injury to a large degree. Traumatic brain injury trials have been typically designed to detect effect sizes of 15-30 percentage point absolute reductions in mortality; for example, if mortality in patients not treated with the new therapy is 50%, can it be reduced to 35% or even 20%. But how realistic is this type of effect size?

It is unlikely that any pharmacological regimen will lead to complete recovery from TBI, particularly injuries classified as moderate to severe, which are those most frequently studied in trials. If we postulate that half of all mortality from TBI would be amenable to protection from any neuroprotective agent, then effect sizes of 20% are attempting to identify 40% of the total pharmacological effect in a single therapy. Many might consider this equivalent to testing a "silver bullet-sized" effect.

Designing trials to detect smaller effect sizes, say in the 5%-10% range for mortality reduction, is likely to be more realistic. Dickinson et al. have shown that only eight of the 203 extant TBI trials could detect 10% or lower reductions in mortality, and none could detect reductions of 6% or lower. It is interesting to note a Cochrane review of six trials of immediate hypothermia, for example, shows a reduction in mortality of 8.1 percentage points based on 230 randomized patients. A total of only 1,040 patients would need to be randomized in a trial to confirm this important degree of benefit from this therapy.

Many TBI trials fail to offer a satisfactory rationale for their chosen effect size. Some appear to be based on extrapolations from animal data, or on availability of patients to a hospital or

trial network, and few use data from phase I or II trials or from other related phase III trials. Only one ongoing trial of corticosteroids for TBI in Europe (the CRASH trial) based its effect size on the results of a meta-analysis of earlier related trials.

Detecting small effects requires trials with more subjects. If it is realistic to expect modest or even small effects, trials should be designed which are large enough to detect them by excluding moderate bias and random error. Small trials have a high probability of type 2 error, which effectively stops research on promising compounds. Based on prior corticosteroid trials, CRASH is designed to detect a 2% reduction in mortality (from 17% to 15%) and may require 10,000 patients per treatment group. It has been suggested that a 2% reduction in mortality offers very little benefit; but, if we consider that over the next decade there will be approximately 5 million head injuries in the United States and 30 million worldwide, a case fatality reduction of 2% would over a decade save 100,000 and 600,000 lives, respectively. The respective lives saved with a 5% reduction in mortality are 0.25 and 1.5 million. Considering lives saved over a decade is justified if this represents the time frame between major therapeutic advances.

How feasible is it to conduct trials of 20,000 subjects? Three trials, of stroke and myocardial infarction, provide sound evidence of feasibility. The Chinese Acute Stroke Trial (CAST) randomized 21,106 patients with suspected ischemic stroke in 413 hospitals to either aspirin or placebo. Absolute 4-week mortality was reduced by 0.6 percentage points (from 3.9% to 3.3%) in the aspirin group, an effect that could prevent 90,000 stroke deaths in China over the next decade. The International Stroke Trial (IST) randomized 19,435 patients after stroke onset in 467 hospitals in 36 countries to a factorial design for heparin, aspirin, both or neither. Absolute mortality was reduced by 0.4 percentage points (from 9.4% to 9.0%) in the aspirin group. This small effect could save 400 lives per 100,000 cases. The Second International Study of Infarct Survival (ISIS 2) randomized 17,187 patients with suspected myocardial infarction in 417 hospitals in 16 countries to a factorial design for streptokinase, aspirin, both or placebo. Absolute 5-week mortality was reduced by 4.0 percentage points with streptokinase, 3.8 with aspirin, and 5.2 with both. Considering the high frequency of myocardial events in the population, ISIS 2 also showed substantial clinical benefit.

It has been argued that a large trial of corticosteroids would divert resources and patients away from other trials. Of half a million patients who sustain head injuries in the United States this year, very few (and sometimes none) are in trials at any one time. Twenty thousand patients in 1 year would represent only 4.0% of available patients. There would appear to be plenty of opportunity for randomizing many more patients to trials than currently occurs.

Large trials require major clinical endpoints, such as survival or major changes in morbidity. One of the most commonly used outcomes from traumatic brain injury is the GOS and, more recently, the extended scale, GOSE. There is substantial evidence that these scales are valid and reliable and their use in large trials is justified. The degree of improvement in GOSE that a trial is designed to detect should be clinically significant and indisputable. There is no value in conducting large trials of debatable clinical importance. The same rationale argues against designing large trials with surrogate clinical endpoints such as disease precursors or physiological and molecular endpoints. Documentation of improvement in anything other than major clinical endpoints is unlikely to influence clinical practice.

Attempts to collect physiological endpoints, or data which may explain mechanisms, is likely to encumber a trial to such a degree that testing the primary clinical hypothesis may be jeopardized. Trials that collect substantial amounts of physiologic data are overly expensive, have lower protocol compliance, take longer to complete and experience decreased accrual, all of which reduce the chance of answering the primary clinical question.

Research into therapeutic mechanisms should be done before a phase III trial or in small random subsets of the larger trial but should not be allowed to jeopardize critically needed phase III investigations. Post trial research may be required to refine knowledge of clinical effects observed in a phase III trial. Few reports of phase III TB1 trials reference phase I or II clinical work, which is where one might expect mechanistic studies to be reported. A rush to phase III trials in the face of quite limited, or even no phase I or II trial data, may explain some of the failure of phase III TBI trials. Another unsatisfactory characteristic of some phase III TBI trials has been basing their rationales on evidence from trials of related conditions, such as stroke, as a substitute for phase II trials.

It may be naive to expect that the dose regimen studied in the first clinical trial of any drug is likely the correct dose. First trials are likely to underdose with respect to efficacy because of concerns about safety with higher doses. With documented safety, dose escalation becomes possible. It is dispiriting to contemplate therapies that may have been abandoned because of lack of efficacy in their first trial, despite good evidence from animal and in-vitro studies of possible therapeutic benefit, which supported further phase III work. Before a phase III trial, drug escalation studies can help determine what a safe and therapeutic dose is likely to be. Reports on phase III TBI trials often do not reference such studies, suggesting they were not done and increasing the risk of phase III failure.

TBI trials have used a broad range of therapeutic windows. Sometimes these windows are long and include patients initiating treatment several days after injury. To preserve homogeneity, it is important in small trials to have narrow windows and these will vary by drug, drug dose, and type of injury. However, very large trials may establish broader windows that can be stratified in analysis to identify particular therapeutic effects. This is an important aspect of the large trials because the opportunity to identify beneficial therapies for some patients may be missed if too narrow a window is prematurely defined.

The majority of TBI trials are limited to severe or moderate injuries. One might argue that these are patients most in need of a successful therapy but this focus may reduce the likelihood of identifying effective therapies. The more severe the injury, the more likely the underlying physiologic response to injury is complex and less amenable to a single therapy. Documentation of benefit in mild injuries may be easier to achieve and may provide clues for treating more severe cases. Trials of mildly injured patients pose some practical problems (they may be more difficult to follow-up), but this hardly seems reason not to study them. Moreover, a majority of TBIs are mild injuries. Thus, while the individual burden is less, the public health burden may be greater. This is particularly so in light of increasing evidence that mild injuries carry more significant and longer lasting disability than heretofore appreciated. Further, it is important to enroll as broad a diagnostic classification of patients in phase III trials as are likely to be treated with the therapy should it prove efficacious. Very large trials can include the entire range of TBI and there are substantial methodological reasons for doing so.

If a therapy leads to an overall shift in improvement across a range of severities, examination of only part of the disease spectrum may be misleading. This bias has been called stage migration or the "Will Rogers Phenomenon." Paradoxically, if a therapy is effective in individual severe TBI patients, it may erroneously appear to make the *group* of severe TBI worse. This is because with therapy, some of the most severe group survive rather than die and they "migrate" to the severe group, thereby worsening average neurological recovery in the surviving severe group. A therapy can reduce mortality and increase the number of patients achieving recovery while appearing to worsen the average neurological scores in survivors. Studying a broad spectrum of injury severity reduces the likelihood of this bias being missed.

In the last 20 years, both public and private organizations have established exclusive networks of trial investigators. The rationale is that only specialized centers have the expertise to conduct trials. This idea may have validity for phase I and II trials, where more detailed investigation is often required, but it has been a flawed concept for the conduct of large phase III trials. The concept errs because it assumes the need to collect large amounts of complex data in phase III trials when this is not the case. Even large networks are too small for trials requiring several hundred centers, and generalizability from exclusive networks may be difficult because they are not representative of hospitals where the majority of patients are cared for. Difficulties surrounding protocol administration and compliance may be underestimated in special networks. The importance of bringing expertise and experience of RCT methodology to many more hospitals than a select few cannot be overemphasized for increasing the infrastructure necessary to conduct large scale trials and, importantly, decreasing the time for newly identified therapies to become more widely adopted.

More successful models establish scientifically credible coordinating centers that can recruit several hundred participating centers using minimal rather than maximal criteria. The CAST (Chinese Acute Stroke Trial), IST (International Stroke Trial), ISIS (International Study of Infarct Survival) and CRASH trials are all examples of this approach. Other examples are the Oxford-Vermont Trials Network with 360 participating newborn intensive care units and the Toronto Maternal, Infant and Reproductive Health Research Unit with over 100 trial centers in 25 countries.

Trials with many thousands of patients cannot afford, and do not need, to collect large amounts of data per patient. Traditional trials typically collect tens of thousands of data points on each patient, although only a fraction of them are likely to be analyzed and contribute to the study's primary clinical conclusion. Collecting these data comes at a high cost. The threats to recruitment, retention and protocol compliance, and sometimes to the integrity of the trial itself, were discussed earlier. The economic cost is also high. A typical NIH-funded trial may cost $10,000 per randomized patient, and some pharmaceutical company trials may cost much more. At these rates a trial of 20,000 subjects would cost $200 million per trial, well beyond the capacity of funding agencies. It is imperative that very large U.S. trials be designed which cost no more than $500-1,000 per patient, or $10-20 million per trial. Widespread adoption of recent proposals to pay research subjects substantial sums based on wage-payment models are of concern and would have a chilling effect on our ability to conduct very large trials.

A reduction in the cost of trials cannot be accomplished without considering more donations of physician and nursing time and the absorption of many trial expenses into patient-care cost. In the context of the present U.S. health care system, this may be an exercise in wishful thinking, and very large trials may continue to be conducted, to a large degree, outside of the United States. Greatly reducing the data demands of a trial should help alleviate some concerns. Unfortunately, large pharmaceutical companies, because of their willingness and ability to pay very high per patient costs in trials, have had a negative influence on some hospitals' agreement to participate in low-cost trials when that opportunity is presented to them.

A further continuing problem in paying for trials in the United States is the unwillingness of health insurers and other third-party payers to reimburse the cost of therapies administered to patients in trials that seek to obtain valid evidence of efficacy and safety. This stands in stark contrast to their payment for many therapies, outside of trials, for which evidence of benefit is either weak or nonexistent. The strategies proposed here for improving the design of phase III trials of TBI suggest that trials should be simpler and much larger. The adoption of very large trials will provide the power and precision to detect small but important improvements in critical clinical endpoints, identification of therapeutic windows and patient subsets where

therapy may be especially beneficial, and the inclusion of more hospitals in important clinical research.

Question: In the CRASH study, the estimate of efficacy is 2%. Was that based on studies of severe and severe/moderate patients? Do you include severe, moderate, and mild injury?

Answer: It is a very conservative estimate. I think there are now 17 steroid trials in the literature that include mostly severe patients (some moderate), and many different steroids (some given two weeks after injury and some given at very low doses). I think the CRASH trial investigators have quite properly estimated the totality of evidence from these trials, found a hypothesis, and will test that hypothesis. The eligibility criteria include up to GCS 15.

Comment: The mortality rate you quoted of 35% applies only to severe head injury, which is only 10% of the total spectrum of injuries. The estimates of lives saved are therefore grossly overestimated.

Comment: That kind of grouping does influence the number of patients. A large number of head injuries in the US are mild. I am not saying that the large trial on all severities could not be done, just that we need to be cautious.

Comment: This is precisely the point. The strategy is to go in, in advance, and look at all of these patients. We (CRASH investigators) planned in advance to stratify on severity, which gives you the option of stopping the trial early in certain groups, if not in all groups. This is a different strategy than limiting this trial to severe head injury patients. If we got an effect in severe we would need another trial for the moderate patients, and years from now, another trial in minor head injury.

Comment: Right now most trials fall between the two extremes. They are not big enough to detect plausible effects, nor are they focused enough to show the treatment effects that Robertson described earlier. If a drug is going to be useful, it has got to do more than work in a few centers where everything is tightly controlled. It has got to work in a large number of centers, so that you will see the same effects when the drug is in common practice that you are seeing in the trial.

Question: Is there any limit on the amount of information that you should collect about patients if you are doing large trials?

Answer: Yes. You cannot conduct these trials and collect the same amount of data that we collect in the usual NIH-sponsored trial. In those cases there are more variables on one patient than there are patients in the trial. Unfortunately, much of this data are never analyzed. I have run trials like that, so I am talking about myself as much as anybody. For a large trial, you will collect limited, but absolutely critical pieces of data. That data have got to be collected on a large number of the patients, and measures have got to be valid. You do *not* want to collect data relating to mechanisms. Data dredging can absolutely kill these trials; you never get around to answering the main question.

Comment: Some data collection is FDA-mandated. Some companies want CT data, and evidence of no harm after drug.

Comment: There are discussions ongoing now in the FDA about what sorts of data requirements there ought to be for various indications. It is not necessary to collect all the information that you think—or a drug company thinks—you need. Certainly, there are discussions that companies can have with the FDA about what is to be collected. When you are talking about things like mechanistic data and imaging studies, there would be no requirement beyond what

good clinical practice involves. When you are talking about laboratory screens for safety, you would not have to get labs on 10,000 patients every 3 weeks. The FDA would require some safety data on every patient that is enrolled in a trial, but perhaps fewer data than you think.

The Long-Term Perspective (John Whyte, M.D., Ph.D.)

I am a rehabilitation researcher, and most of my work concerns subjects who are the failures of neuroprotective research. It has been intriguing for me to see how many of the problems in acute trials are also found in rehabilitation, though sometimes in somewhat different terms or viewpoints. How do the outcomes that result from clinical trials, or from the natural course of events, play out over time, and what implications do they have for research design?

The hierarchy of outcome concepts, described by the World Health Organization's ICIDH (International Classification of Impairment, Disability and Handicap)-2 or recently published by the Institute of Medicine, is critical to an understanding of clinical trials in brain injury and to selecting the appropriate outcome measures for such trials. In this lexicon, pathology refers to lesion volumes, degree of diffuse axonal injury, ICP and so on— that is, damage at a cellular and tissue level. Impairment refers to the dysfunction of organ systems that might be manifested by such things as reduction in level of consciousness, paralysis, or memory deficits. Disability (or activity, depending on the system of nomenclature) resides in a person rather than an organ system, and includes functional abilities, self-care or mobility, and so on, that is, the ability to perform the normal tasks of daily life. Handicap (or participation) exists at the interface between a person and an environment or society and reflects an individual's ability to assume expected social roles. So, a person's ability to be employed is a product of their own skills and abilities in interaction with factors such as the types of jobs that are needed in the community and the accessibility of transportation.

Quality of life represents a global assessment of an aggregate of many features taken together. In theory, one could measure the outcomes of clinical trials at any of these levels, but there are clearly tradeoffs in doing so. The question is then how do you select among those various outcomes for research purposes? The "microoutcomes," that is, the ones closest to pathology, show the greatest sensitivity to biomedical interventions. All of the treatments and agents that have been discussed at this workshop are designed to influence most directly tissue pathology. Thus, the linkage between successful intervention at this level and reduced neuropathology will be tightest, and dependent measures of neuropathology will be most sensitive to the effects of such agents. However, the outcomes that people and society are most interested in are at the other end of the spectrum—the more "macrooutcomes" of disability/activity and handicap/participation, which reflect the ability to return to independent living and quality of life. Those "macrooutcomes" are more tenuously related to the early biomedical factors. A person who has a traumatic brain injury may have other coexisting conditions (e.g., osteoarthritis, diabetic neuropathy) or other concurrent trauma (e.g., traumatic limb amputation) that may influence outcome. These multiple pathologies may collectively lead to a variety of impairments (e.g., paralysis, balance deficits, reduced proprioception). Any given neuropathologic lesion may produce more than one impairment, depending on its size and location. Thus, there is no simple one-to-one correspondence between TBI and any particular set of impairments, though some are certainly more common than others.

Moreover, several impairments together may impinge on a functional ability. For example, stair-climbing ability is related to muscle tone, balance, strength, and pro-prioception. If neuroprotective agents in the above example reduce paralysis and normalize muscle tone, this will not erase the effects of diabetic neuropathy and limb loss on balance and proprioception. Thus, the "successfully treated patient" (in the sense of early neurosurgical intervention) may still show a stair-climbing disability. Obviously, stair-climbing may be a minor contributor to

return to work (a handicap/participation outcome), but that depends on whether one is an office worker or a letter carrier.

In summary, while one may successfully reduce neuropathology through acute neuroprotective intervention, tracing that effect up this conceptual hierarchy to daily activities and societal participation, over months or years, inevitably results in the loss of considerable explanatory power. Even highly successful biological treatments will account for only some of the variance in these multifactorial biopsychosocial outcomes because they affect only some of their causative mechanisms. At the most simplistic level, if one wishes to perform clinical trials that involve "macrooutcomes" as the dependent variables (and the GOS is such a measure), this will always require a larger sample size to account for the noise produced by the many coexisting causal factors operating at these "macro" levels. Another issue to consider is that recovery in traumatic brain injury continues for a prolonged period of time. We do not really know how long that is, and the duration depends on how we define recovery. For example, as long as the person retains some consciousness and is able to learn, at some rate, skill development and adaptation to disability or environmental challenges will continue over years, if not indefinitely. Even at a more biological level, we have a small number of individuals who have been vegetative or minimally conscious for 2 years who then have regained consciousness or had substantial increases in function. When following patients for prolonged periods, as physiatrists often do, one does see some ongoing change, though the rate slows over time.

In clinical trials, one may want to measure the level of a "final outcome," but when is an outcome final? At what point in time should one measure? If one wishes to measure outcomes relatively early (e.g., 6 months after injury), and chooses something like neuropsychological performance, there will be whole populations of people who cannot be validly assessed. The "outcome" becomes very complicated to analyze. Another approach is to look at the time until some standard outcome is reached. This is essentially a "survival analysis" methodology, where one can ask, "What is the time until consciousness or a certain neuropsychological profile is obtained?" There are methods of censoring data for subjects who have not reached that endpoint by the end of the observation interval. Typically, this design has a lower statistical power than a fixed point in time for each subject.

In my own research, early biological variables tend to account for less of the outcome variance over time. For example, evoked potential testing can predict outcome when done within the first 2 or 3 days after injury. We were interested in seeing whether results of somatosensory evoked potentials (SEPs) obtained months postinjury, predicted further recovery in a severely injured group of subjects. While there was a statistically significant relationship between late and further recovery, behavioral measures of function were far more strongly predictive of further functional change than were SEPs. Similarly, in our work on predicting and influencing recovery of vegetative and minimally conscious patients, we find that among patients who are still at this level at least 4 weeks postinjury, early GCS scores and biological measures offer little predictive value. Current functional measures and short-term change in those functional measures are strongly predictive, accounting for about 70% of the variance in subsequent recovery of consciousness. Thus the importance of biological variables as outcome predictors or stratification tools appears to diminish over time.

There is an additional sampling issue that is faced, particularly in performing interventional research at later time points. In rehabilitation research, we see that the patients who are treated in rehabilitation settings are a biased sample being neither "too good" nor "too bad." Studies that seek to enroll patients at that point in time or to follow patient outcome are going to seriously misrepresent the total spectrum of TBI patients.

What can we recommend from some of these perspectives? As a physiatrist researcher, I may risk accusations of heresy, but I believe that it is a mistake to use more "macrooutcome" variables for research on early stages of neuroprotection after TBI. As mentioned, such measures will be relatively insensitive to treatment-related changes taking place at the biological level. A more reasonable approach is to focus on fairly "microoutcomes," things close to the pathology level, during the initial phases of research. The ideal scenario would be, for example, to study the impact of a neuroprotective agent on proximate pathologic variables such as reduction of ICP, improvements in brain metabolism, and time until consciousness is regained. Very likely, such research would identify subgroups for which the treatment was highly effective and other subgroups where it was not, given the heterogeneity of injury mechanisms in TBI. Surely, it would only make sense to expect improved mobility and self-care in the subgroup that actually had reduced neuropathology. Thus, a second wave of research could focus on the effects of this same neuroprotective treatment, but now in a more restricted subject group that is known to benefit at the level of pathology. This same logic can be followed up the conceptual hierarchy, rather than expecting a discernable effect of the neuroprotective agent on such outcomes as employment in a highly mixed population. The argument here is rather than doing a huge trial to detect the needle in the haystack, do earlier phases of research to identify ways to predict the subgroups whose pathology will be ameliorated. Then use those as selection approaches for larger studies that extend findings to ask whether function is improved, and move forward step by step.

Another thing that is evident from research in our laboratory is the importance of descriptive outcome modeling in guiding clinical trials. Today, speakers have addressed the utility of surrogate markers, but surrogate markers are useful only if their relationship to the true outcome of interest is well understood. In our work on prolonged unconsciousness, we have invested a lot of energy in identifying the predictors of our outcomes of interest, so that we can then look with greater confidence at factors that influence those predictors, in the hopes that they will also influence the outcomes of interest. If there is little relationship between a particular early factor and a later outcome, then it does not make sense to study that early factor.

In summary, clinical trials of neuroprotective agents need to pay careful attention to the conceptual level of the dependent measures chosen, and should attempt to carefully and sequentially trace the impact of treatment from the level where the treatment acts to the levels of desired impact. Tracing these effects across levels and through time is a painstaking process and each step risks the loss of explanatory power.

Question: I agree with the notion of focusing attention on early outcome of biomedical interventions, but there is danger in not looking long-term in those situations. For example, in coronary artery bypass, if you select the early outcome of graft patency, you can conclude that the procedure is efficacious for all subgroups of patients, whereas if you consider longer-term outcomes, such as death or disease progression, you find very different answers.

Answer: That is certainly true, but your question assumes that any one of these stages works alone, and I am arguing for a programmatic sequence of research. Take the flip side of your example: in the initial study we find that coronary artery bypass grafting does *not* result in patent vessels for some patient subgroups. Will we want to go on and study functional outcome in those groups, when the basic mechanism of treatment has not been obtained? Suppose we discover that a particular type of vessel stays patent when grafted, while another type tends to reocclude. We may then want to design our functional outcome study around the people receiving the biologically successful graft. We might postpone a functionally oriented study on the other subgroup until we have identified a surgical procedure that is equally successful in retaining graft patency.

PRACTICAL ISSUES

Staffing for a Clinical Trial (Nancy Temkin, Ph.D.)

I shudder to hear about trials where centers are entering a total of five patients. It is unlikely that such a trial will have tight quality control and good compliance. For large trials, even at a single center, if the study personnel are not on top of the patient enrollment and follow-up constantly, the entire trial is lost. Problem areas are everywhere, from enrolling patients in a timely fashion through ensuring that patients return for testing. The TBI population is difficult to work with. Many are "risk-taking" young men, and coming back for follow-up is not their highest priority. Adherence to the protocol schedule is difficult, and only gets more difficult over time; a 2-year follow-up can be a daunting task. One of my favorite stories illustrates the problems study coordinators face. An individual came back for the 2-year follow-up, and we asked if he wanted to know what treatment he was on. He said, "Yes." We broke the blind and told him he was on placebo and he said, "and to think my friends were paying good money for those pills."

Our Dilantin study was a 1-year treatment, and it was an interesting job keeping compliance at an acceptable level. People with severe head injuries often have other systems injured. Their metabolism changes over time, and to keep a consistent drug level, we had to keep manipulating the dosage. It takes a large and really dedicated group to pull off one of these trials; our group has 50 people. The ones who make the trial go are the nurses and the outcome examiners. They interact closely with the patients, and must bond with them to get the high follow-up rates. Our study nurse describes herself as a pit bull with good social skills.

Data Collection in Traumatic Brain Injury Trials (Lorraine Yurkewicz, Ph.D.)

Due to the medical complexity and severity of acute TBI, and the inherent difficulty in demonstrating the efficacy and safety of new therapies, there is a tendency to collect as much information as possible in clinical trials. This additional information has often proven useful retrospectively, in determining some of the contributing factors (i.e., patient heterogeneity, differences in medical management) in failed clinical trials. Despite its usefulness, the volume of data collected in a TBI clinical trial can become enormously resource-intensive, both for the research staff and for the sponsor. It is therefore important to determine which data are critical for the assessment of key efficacy and safety parameters for scientific validity and regulatory approval. The resource burden for any additional data collected must be carefully weighed against its potential benefit.

I am currently conducting the Pfizer study in severe head trauma. I would like to talk about some very practical issues surrounding data collection that I have been dealing with for the last two years.

- Patient population. There are complex issues of patient entry criteria for TBI. It is important to decide whether or not to include minimum GCS scores of 3 or 4. The baseline criteria between the two treatment groups (drug vs. placebo) need to be balanced, and there may be ways of stratifying at baseline to ensure this. Which types of TBI to include, as evidenced by CT scan, is controversial; yet it is critical to determine beforehand what patient population will be studied. Should gunshot wounds be included? Open and closed head injuries? The precise patient population to be included will also depend on the drug's mechanism of action.

- A consistent treatment paradigm. One of the biggest practical obstacles that we must deal with is time from initial injury to dosing. It is important to administer the drug as soon after injury as possible, but we have the practical constraints of obtaining written informed consent. There needs to be a balance between making the window

as short as possible for maximum drug effect, but long enough to obtain consent and enter patients into the study.

- Optimal drug dosage and duration. Rely upon information gained from early phase trials and pre-clinical studies. Be practical with the regimen to reflect the necessities of treating severely injured patients.

- Endpoints. One must define the primary endpoint in terms of efficacy. There is some disagreement as to what is the best primary endpoint. Until now, most people have used the dichotomized GOS, and I know Pfizer is using it in our study. According to Dr. Choi, it seems to be the most sensitive outcome scale in terms of power, but maybe there are other endpoints that will prove to be more sensitive and will better reflect improved cognition. It would be very useful to find a physiological endpoint that could be used to show drug efficacy earlier than at 6 months, and that predicts long-term outcome.

From a sponsor's point of view, every data point collected involves resources: research and hospital staff, a study coordinator, a field monitor to come out and verify every data point against the source documents, a data manager, a staff for data entry, a statistician, and a programmer. In our study we have attempted to streamline our casebook, yet we have about 50 pages and at least 3,000 data points per patient. Someone has calculated that it costs a company $100 for each data query. For product approval there must also be proof of safety, and this requires reporting all adverse events, whether or not they are considered to be related to the study drug. These reports are resource-intensive. To understand and interpret the results and to determine the safety profile of a new agent, one needs to collect a lot of data regarding a multitude of medical issues for each patient. In severe head trauma there are many adverse events. Issues regarding the collection of laboratory data for safety must be decided early on. Does one collect all of the data or only selected parameters? What types of treatments are involved other than the study drug? Should there be special chemistries, EKG profiles, or CT scans collected during the study. Will these demands affect patient management and standardization across centers? In our ongoing study we have spent a great deal of energy on standardization of patient management. These issues are critical in terms of outcome.

One may choose to evaluate secondary measures or exploratory endpoints in a study. For example, collecting data for subpopulation analyses, including baseline characteristics and patient management, may determine which parameters are most critical in influencing outcome. Planning resources for these analyses at the start of the trial will avoid surprises later and will help in focusing the efforts of the investigators.

In addition to the primary efficacy endpoint (such as the GOS), it may be useful to include a cognitive neuropsychological scale. It is important to avoid the danger of secondary scales being too lengthy. This would represent a burden, both for the research staff and for the patient. Complex outcome instruments may make the patients less willing or able to cooperate. Another parameter to consider including would be a measure of the economic risk/benefit ratio for study drug treatment. This may become more important in the future, and it is a very difficult measure. One option is to compare treatment groups with respect to length of hospital stay or the time needed in acute care. What may become more important eventually is to examine the overall cost of illness. This is quite complicated, because billing practices vary from one hospital to another.

Additional data that may be useful to collect are found in the medical history. One thing that is very important to know is whether or not a patient has had a previous brain injury; that certainly might affect outcome. It would also be important to know if a patient had any kind of cognitive impairment prior to the TBI, because that could confound the interpretation of results. Two items that are necessary to collect but that are also very resource-intensive are

concomitant medications and concomitant treatment. Without this information it is difficult to interpret any kind of abnormalities or adverse events.

The kind of information that may be useful to collect in terms of standardization across sites and patient management are pupillary status, GCS, vital signs, and ICP. In the acute care period, one may be collecting ICP and vital signs every hour, for several days, resulting in a huge number of data points. The intensity of treatment intervention required to reduce ICP may be very important in comparing treatment groups. In fact, if the research drug is working and the trauma centers are practicing aggressive patient management, the only difference between treatment groups may be the amount or intensity of intervention. One would expect the placebo group to require more aggressive treatment than the active drug group.

Comment: There are ways in which data collection can be reduced. In studies where the treatment is very acute, but outcome is assessed at 6 months, the FDA would not ordinarily require collection of every data point throughout the time of the assessment, as they might in a study where drug is given chronically for six months. At the FDA, we are toying with the idea of requiring "labs" be collected only for perhaps five half-lives of a drug. The reality at FDA is that the sponsors (drug companies) come to us and we review their protocol; if it is reasonable, we say it is possible to begin the trial. Companies rarely ask us if they can do something different; we would consider other reasonable safety protocols. It is not necessarily the case that every single concomitant medication a patient takes must be recorded and examined. There is some room for discussion. People seem to love to assess and analyze secondary endpoints. Secondary endpoints can present problems, and are not usually considered when the FDA decides whether or not a drug is effective. When the primary outcome is negative, people invested in the trial may try to resurrect it by saying that a particular secondary outcome is positive.

Quality Assurance in Intensive Care Unit Data (Anthony Marmarou, Ph.D.)

We had the unique opportunity to explore variation in ICU management over the last few years. The advent of the Guidelines sponsored by the AANS (American Association of Neurological Surgeons) and the Brain Trauma Foundation gave us a "before and after" from which we could compare. In March 1996, the Guidelines were sent to all the surgeons in the AANS registry and were published in the *Journal of Neurotrauma* in November 1996. We considered that we had two data collection periods: before Guidelines (1995-96) and after Guidelines (1997-98), and we assessed the degree to which the recommended procedures were being followed.

ABIC treatment centers collected CT scans and acute care parameters of ICP and TIL for 6 days. Copies of the medical records were faxed to the consortium center where we read the CT scans and reviewed patient management. Reports, tables, and graphs were sent back to the study center to show them problems or deviations from protocol. Our hope was that center staff would attend to these recommendations and improve their management.

Our study population was 326 patients from two prospectively randomizing clinical trials at 104 centers, which represents a reasonable sample of what was happening in TBI trials the United States at that time. The patient ages ranged from 18 to 65 years. The inclusion criteria for both trials were "one reactive pupil and abnormal CT" and the selected populations did not include the most severe end of the spectrum (GCS-3). The study investigator received graphed reports that compiled the first 120 h of the intensive care record: blood pressure, ICP with threshold for treatment, cerebral perfusion pressure with a threshold (listed as an option by the guidelines) to maintain above 70 mm Hg, and TIL. In the reports we can see individual treatment profiles: as a result of high ICP, there is aggressive therapy and significant reduction in perfusion pressure. We can also pinpoint possible protocol deviations or failure to comply with the Guidelines. In one example, we saw an elevated blood pressure with the TIL consisting

of sedation, frequent mannitol, and hyperventilation, but no ICP monitoring. What was guiding this therapy? It looked as if the intensive care staff was using a shotgun approach. This finding called for our contacting the center and discussing the reasons for the deviation.

If you look at the compliance with Guidelines from April 1996 to 1998, you see that only 58% of patients were managed according to suggested practices. If you look at the reasons for noncompliance, most of them dealt with sustained hyperventilation (>4 h). In some of the cases there was no need for sustained hyperventilation because there was no ICP problem. The other deviation was the absence of pressor use with CPP <70. Surprisingly, in 14% of centers there was no ICP monitoring.

When we looked overall to see deviations from the Guidelines after publication in 1996, it looked as if there was a reduction in those deviations that were considered moderate; however, the number of severe deviations increased! There was marked reduction in instances of insufficient CPP, and the ABIC centers were using more pressors; however, there was a *decrease* in ICP monitoring, up to 27% of centers. I could not believe this, but there it is! There did not seem to be an improvement in treatment after publication of the Guidelines. What are reasons for this noncompliance 18mos after distribution of this new information? This result is rather discouraging for those who plan and direct clinical trials. Noncompliant centers had more difficulty in trying to control ICP, and the percent time for CPP < 70 mm Hg was very high in these centers. If you look at the effect of compliance on 6 months GOS, 44% of patients from compliant centers had good outcome versus 34% from noncompliant centers. This was the first evidence that adherence to the Guidelines can have an effect on outcome. Another important point in running clinical trials is center compliance. In our analysis, centers that enrolled fewer than five patients had a 45% compliance rating, whereas centers that enrolled more than five patients had a 65% compliance rating.

We have a significant number of deviations from the recommended Guidelines among ABIC centers. There did not seem to be any improvement in the overall frequency of deviations in patient management after Guidelines were publicized. The greatest number of deviations was associated with centers accruing fewer than five patients, which heightens our concern about center differences in management of severely head injured patients for the conduct of trials. Adherence to compliance guidelines appeared to have a favorable impact on outcome, although more data are needed to confirm that this is significant. Our recommendation for trials is that an in-depth clinical review be an essential component of the study. In a clinical trial designed to detect small effects, suboptional management would completely invalidate the outcome. An educational process on the importance of following the Guidelines must be ongoing and must extend to neurosurgeons, intensivists, trauma specialists, and any staff who are involved in the acute treatment of severe brain injury.

Question: It is a little bit disturbing to me that your study and the Guidelines focused on CPP < 70 mm Hg. We tend to forget that cerebral perfusion pressure is a derived value that depends on several factors.

Answer: We are now looking at the effect of CPP on outcome, and there is a significant effect, albeit that the initial ICP is extremely important. I do not believe these data speak to the "appropriateness" of where to intervene. The threshold for CPP was set at 70 mm Hg in the Guidelines. Is 50-60 mm Hg more appropriate? We are not sure. We do have data on the percent time spent at these different levels, and that issue should be sorted out.

Comment: In the phase III hypothermia trial we found that CPP of 60 mm Hg or less was the critical number, with really no effect at 70 mm Hg. No one has found much effect of CPP > 70 mm Hg. If you target CPP = 60 mm Hg for the "alarm to sound," the value will be less than

this for some of the time; it looks to me that you need to aim for about 5-6 mm Hg over the critical variable.

Answer: Another piece of important information was that even a single event of CPP < 40 mm Hg was devastating. So, we are in search of critical thresholds— whether percent time, or frequency, or level.

Question: You criticized sustained hyperventilation as a serious violation of the Guidelines. But, the ABIC Guidelines include hyperventilation as an option; not perhaps standard treatment, but an option. Were the participating centers all neurosurgery units or also general trauma?

Answer: These were all ABIC neurosurgical centers. The guideline was not to use prophylactic hyperventilation in the absence of elevated ICP.

Consent Issues in Acute Care Trials (Charles Wade, Ph.D.)

Most of our work has been looking at hypotensive patients, and in that group, traumatic injury to the brain is a subpopulation. I am going to use some abbreviations: the IRB and the legally authorized representative (LAR). We must have informed consent from the subject or the subject's LAR in order to carry out experimental studies. Defining that individual can get very difficult: a wife, a family member, or parent. The LAR must be defined before you begin, and it is different in various states. Institutions can impose requirements in addition to the state law. Another issue that lawyers have a tough time with is, "Can somebody who has traumatic, physical injuries or their representative at that time provide informed consent?" Does the medical situation that we are working in constitute an inappropriate place for an individual to give informed consent? Consider the situation for a family member of a patient who has been in a traffic accident. It is late at night; there are helicopters and paramedics, emergency room personnel, resuscitation, and severe injury. These activites were designed to save lives and optimize outcome, but can you expect an LAR give true *informed* consent in that environment? There are lawyers who say no.

We face a disease with grave consequences and no proven pharmacological therapies. Patients with acute traumatic injuries surrender control over early care decisions—they have entered a system of transport, ER, ICU, OR. Often they are not aware of or do not understand the procedures. They are in a mode of passive acceptance and loss of autonomy. Once in the system, they tend to roll with it, even after they get out of surgery, or acute care. This state of mind is referred to as "institutional transference." What should investigators do?

1. Try to get informed consent by the patient when feasible. Remember that the patient always has the right to decline, so you have to give them that option. Even if you think they are drunk, dizzy, unaware of surroundings. If they decline, they decline.

2. Propose consent by family or the LAR when feasible. In the TBI trials, you are usually trying to get consent within the first six hours. I am trying to get consent in the field, pre-hospital, to administer fluids before the ER, or immediately upon admission.

3. Have a scripted, abbreviated consent form. I went back and looked at some work we had done in 1983. The consent form in that environment was three pages long, in 14-point type. Our consent form today is 14 pages long, in 12-point type, single-spaced, and there are descriptions in it where I wondered what was going on. Consent is supposed to be in layman's terms. Sponsors have gone to an abbreviated consent form for this emergency environment. The patient or LAR can read one clear page very rapidly, sign it and go forward. The lengthy document comes later.

4. Independent approval or the use of a consent exception by a second physician prior to enrollment. Such an arrangement should be in place prior to the beginning of the trial.

5. Repeat the consent to continue when an exception to prior consent has been utilized. Throughout our work we have learned to periodically ask the patient if they want to stay enrolled in any particular study. This procedure is useful especially when we are following subjects for a long period of time. We give them an option throughout the whole study to withdraw if they wish. That vigilance helps in terms of the ethical issues involved. There are reports on studies done on patients in the emergency room. In one study (not TBI) prospective consent was given by 6% of the patients. Enrollment criteria required a blood pressure <90 mm Hg and major traumatic injuries. The questions in using prospective consent are, "Is the patient with major traumatic injury competent? Is the family member competent?" In this study, independent approval for a consent exception (waiver of consent) was given in almost 94% of cases. They used a second physician to independently evaluate the patient and determine inability to consent. For research done in the field, i.e., emergency or prehospital care, the institution has to take responsibility for this consent approval process. Getting an institution to do that can be a problem.

Consent to continue is used in 89% of the patients. Usually, the investigators ask a family member within 24 h and then later talk with the patients. The average time for consent to continue obtained directly from patients was 13 days later. I want to emphasize that we keep asking the patient over and over to affirm that they want to participate. We get the family members and the patient to play a role in continuing consent and get them to sign off on it.

Continued IRB review has to be implemented in this particular environment. It is difficult for many IRBs to assign someone to oversee emergency research because it is a major activity. Another requirement for using waived consent concerns community consultation and public disclosure. There has been a real issue of interpretation here. What are we to do and how do we know it is appropriate? I will give you an example. For a study in an East Coast state on trauma patients in the ER, the investigators went to church groups, radio talk shows, and such activities. When the study started, the typical trauma patients were young males, and this population did not typically appear in church or listen to talk radio. Do you earmark this "information" effort towards your patient population? Or to society as a whole? It was suggested in one case that investigators go to the gangs, and discuss the proposed research. That is how extreme this is getting.

Our field needs to discuss the objective in emergency medicine research. Should it be potential improvement in survival? We have the possibility of getting a positive result, and survival is the bottom line. That is what has been accepted among the trauma community and acute care community. In most of the studies we are discussing, we would administer the standard of care and add to it. We are looking for something that we believe is of benefit in addition to standard care. In our comparison between hypertonic saline and standard of care, we saw that with the group of subjects with GCS<8 there was an increase in survival. The treatment significantly elevates blood pressure by expanding blood volume. That increase persists into the ER, and we see an increase in survival. There was only one group that showed benefit when we looked at it in terms of using an AIS (Abbreviated Injury Scale) of >4. There is a problem: When is that determination made? It is made a couple of days later, when I have the CT scan, the autopsy report, and the whole record. Is the determination of AIS biased? We looked at this when we did the meta-analysis; we used 1,080 patients, and we had 224 classified as having head injury with AIS >4. When we used the GCS <8, we had 52% of the patients who actually had head injury.

To convince an IRB to allow very early studies, we need adequate basic research. In terms of hypertonic saline/dextran, we had done over 400 animal studies prior to clinical studies. There must be adequate preclinical research, and it has to have been evaluated. We must take published data to set a reference point or a standard. A good IRB will look for that.

Question: It is my understanding that about 4 years ago the staff at FDA and OPRR determined that "deferred consent" is an inappropriate term. (The Office for Protection from Research Risks no longer exists. Recent reorganization has created the Office of Human Research Protections within the Department of Health and Human Services.)

Answer: That is correct. The terminology is "waived consent." As I said before, you cannot give up getting consent. All you can do is waive consent for a period of time.

Changing Clinical Practice in the Real World (Jamshid Ghajar, M.D., Ph.D.)

There is a gap between science and clinical practice. I think we can close the gap, but not by lecturing or going to conferences. It must be done using databases to track patient assessment and treatment. I think Dr. Marmarou's talk addressed two things. One was compliance with the guidelines, and the other was the difficulty that ABIC had in convincing the trauma centers to follow their protocol. We are all facing the question, "How do you make trauma centers follow protocol?"

In 1991, the Brain Trauma Foundation (BTF) surveyed about a quarter of the 1,000 trauma centers in the US. We found that 26% were not monitoring ICP routinely yet were using severe hyperventilation and steroids. The BTF and the AANS developed the Guidelines hoping that, by bringing together all the scientific knowledge and distributing it, treatment practice would improve. We repeated the survey after publication of the Guidelines. We had three nurses working full-time who contacted the American Hospital Association and every single state department of health, and drew up a list of trauma centers in the US. We had 90% participation in the survey. For 50% of the trauma centers, the TBI patients came in, and then were transferred at a later time to another trauma center. That is an important point, because there are data to show that transferring patients can increase mortality. I think we can state that about 500 trauma centers in the United States are really taking care of all the head injury. We then sent questionnaires to those centers, with 35 questions relating to how many patients were seen every month? Who heads the ICU? In preparing the questionnaire we went through the Guidelines: prehospital issues, trauma systems, indications for ICP monitoring, the threshold technology, thresholds for CPP, hyperventilation, mannitol, barbiturates—all ICU issues except nutrition were addressed. The forms also included case scenarios of how the center would manage a head injury patient.

Our scoring scheme was this. Full compliance: the center had ICP monitoring routinely in 75% or more of qualified patients, followed the ICP treatment threshold, the technology, hyperventilation, mannitol, and steroids. Partial compliance: the center monitored ICP routinely as above. Minimal compliance: no routine ICP monitoring. Results indicated that two-thirds of the trauma centers were out of compliance. Only about 15% were in full compliance and about 18% in partial compliance.

If you look at each question of the survey and then look at level I (the highest designation), level II, and level III you see a significant difference in the level I trauma centers compared to level II and level III in terms of ICP monitoring. Level I trauma centers had a much higher incidence of ICP monitoring compared to level II and level III. Mannitol and barbiturates were used significantly more in the level I trauma centers. Steroids were used significantly less in the level I trauma centers. Level I centers were clearly much more "Guidelines-compliant" compared to level II and level III. The level I trauma centers saw much higher volumes of TBI

(>15 cases/month) than the level II or level III. These level I centers saw in the range of four to 14 patients with severe head injuries each month. It seems that high volume and being a level I trauma center led to more compliance with the Guidelines, than did a low patient volume or level II or level III status.

Using just routine ICP monitoring (>75% of qualified patients) and technology (ventriculostomy) as indicators, we had 26% compliance in 1991. The Guidelines were published in 1995, and ICP compliance went up to 33%. That looks optimistic, and if we wait another 50 years we might see full compliance. I think what this report and what ABIC's data are saying is that we cannot just publish a book and have things work. How can we reeducate physicians and medical personnel? This is a big question for all areas of medicine, and for TBI clinical trials.

We are trying to develop guidelines, implement them, and conduct clinical research in order to create a continuous feedback loop of science and practice. The really tough part of having practice guidelines is implementing them. The BTF was fortunate through a grant from the Soros Open Society Institute to develop an educational program in Eastern Europe. The effort created an Internet database that included the CT scans, so we had central CT reading in this project. The interactive Internet database to assess the efficacy of the Guidelines demonstrates improvements in patient outcome and allows online education of physicians. The program is now ongoing in seven countries in Eastern Europe: the Baltics, Czech Republic, Hungary, Slovakia, and Croatia, in the major trauma centers. In each center there is a neurosurgeon in charge of the program and data entry.

There are about 1,000 patients in the database, and it includes everything from prehospital vital signs through 6-month outcome. We combined the databasing effort with an annual conference for the centers, and many of the authors of the Guidelines attend and give lectures or hands-on workshops. CT scans were transmitted back to use in New York over the Internet, and two neuroradiologists at Cornell read and graded the scans.

Results from 850 patients include the following: Motor vehicle accidents were the leading cause of injury. Age range was 3-60 years. Forty-one percent of the patients were transferred. GCS was 3-5, noted after resuscitation in 66% of patients. There was ICP monitoring in 58% patients. In the first three days, there was a very low mortality when ICP was <25 mm Hg for the entire time or was >25 mm Hg for <40% of the time. There was a high mortality in the group that was not monitored and in patients with ICP > 25 mm Hg for >40% of the time. ICP data are clearly important in predicting outcome.

Our data indicate that CPP levels above 60 are not strongly prognostic. What is prognostic is mean arterial blood pressure. We used low MAP as 85 mm Hg, and we found that for high ICP (>25 mm Hg)/low MAP, there is 26-fold increase in mortality at 2 weeks. For patients that show low ICP/low MAP, there is a sevenfold increase in mortality. This data indicated that we should not be looking at CPP as much as ICP and MAP individually, and really look for an adequate level of MAP.

Was our education effort successful at our 10 centers? We looked at ICP monitoring as an indicator of following the Guidelines, and when we began there was monitoring in fewer than 10% of patients. Just by walking in and shaking the physicians' hands, we saw that monitoring went up to 40% of patients. The investigators entering information on their patients knew that it would be in a database, and they wanted to do the right thing. But you can see that over the years ICP monitoring has continued to climb. I think that "databasing" makes a big difference in compliance with the Guidelines.

This original database, called TBIS, was very expensive. There now is a new database, TBI-trac, which is very short and can be done in 30 min on a patient after 2 weeks. New York State is funding this grant, and there are have five trauma centers participating now. They are online through the Internet starting June 1, 2000 and they will enter all the acute data from patients from the moment of the accident through two weeks. At the end of each month, centers will get a report about their patients, showing which patients were "guidelines compliant" and which not. When they enter a patient's data the program will immediately show them online whether the patient is compliant with the Guidelines or not. We include whether the particular guidelines compliance is at the option, a standard, or guideline level.

The New York State project also has an educational materials website (http://TBI_trac.edu) that has the Guidelines on it. This site allows medical personnel to see videos, and to access every abstract and evidentiary table that was used in developing the Guidelines. The BTF has also developed prehospital guidelines for severe head injury under a grant from the U.S. Department of Transportation. This education effort is spreading very quickly, because the 800,000 EMS people are taught by this agency. There is one office in Washington, D.C. responsible for education, and once we had their approval, we moved rapidly. These guidelines are detailed and include a trainer manual, a student manual, and films. We had pilot sites in Anchorage, Alaska; Navajo Nation, Arizona; Galveston, Texas; Birmingham, Alabama; and Washington, D.C. The project is now moving into several other states.

In summary, for physician reeducation, I think you have to do databasing and you have to do active education.

Comment: You have an opportunity to do some randomization here. Perhaps randomize hospitals to different procedures within the guidelines. If you do trials in that area you would be contributing to some of the general knowledge on how to implement guidelines in any specialty. I believe that neonatologists are doing this type of thing. I would encourage you to try and think of an experimental design.

The Perspective from the Food and Drug Administration (Russell Katz, M.D.)

My division at FDA (Division of Neuropharmacological Drug Products) has not interacted regularly with the neurosurgical community in general, or the head trauma experts in particular. I do not know exactly why that is, but one reason is that meetings like this are not held very often. I think it is very important that we get a chance to talk about the FDA standards and what the rules are for drug approval. I know that your perceptions of them are largely through your interactions with the pharmaceutical industry, and it may be useful for you to hear from us. There might be a number of myths out there, and I will explode some, confirm some, and perhaps create some new ones.

There is a statutory standard for determining effectiveness. The Federal Food, Drug and Cosmetic Act says that in order to approve a drug there must be substantial evidence that the drug will have the intended effect. What exactly does "substantial evidence" mean? Everything with regard to effectiveness is related to the labeling. What claim is the sponsor proposing, and does the evidence support that claim? It is a vague standard because claims vary from drug group to drug group, within a drug group, and even within a particular therapeutic area.

You have identified many questions for the design of TBI trials: large simple trials or specific trials, the entire range of head injury patients or specific subjects, Glasgow Outcome Scale or neuropsychological testing? The reality is that any of those options are potentially appropriate. There is no specification from the FDA about what type of trial needs to be conducted. All of those things are potential areas for negotiation. For example, a drug could be approved just for

severe head injury, or for head injury in a very global sense, covering a range of severities. There is no limitation from the point of view of the statute (the law) as to what can be approved.

Generalizability of findings. Generally speaking, the FDA is not as concerned about a couple of things that sponsors and investigators talk a lot about. One is the generalizability of the treatment to the larger community of patients. We are looking for trials that show a drug effect, that is, a difference between the treatment and control group. The drug does not have to be for everybody who has that disease; it could be limited to a very restricted subgroup. Remember that the samples of patients in clinical trials are not random samples of the population that has the condition. They are what my predecessor used to call "samples of convenience." The FDA wishes to ensure only that the trials are designed in a manner that the effect we see is a *bona fide* effect, even though the patients in a TBI trial do not represent all patients with TBI.

Size of treatment effect. We also do not care so much about the size of the treatment effect. We are interested in trials that can demonstrate a difference between the drug and the control, a proof of principle. You must show that the drug has a pharmacological effect on an outcome that means something clinically, in an appropriate sample. In the world of epilepsy, we used to look at 50% seizure response. In other words, a patient was a responder if there was a .50% decrease in his/her seizures. But what about a 48% reduction? It is hard to call that person a failure. We at FDA do not set a standard for the size of the treatment effect. As long as a *bona fide* difference has been demonstrated between the drug and the control, that is good enough.

"Substantial evidence". The law also specifies that you need adequate and well-controlled investigations, including clinical investigations that demonstrate that the drug will have the effect it is represented to have. Again, consider the proposed labeling. "Clinical" has been determined to mean human, so you have must do studies in humans. "Investigations" is written in plural, and we have invariably interpreted that to mean that you need at least two adequate and well-controlled trials. What is an adequate and well-controlled trial? The regulations written to interpret this say there are about five different types, and they can overlap: (1) placebo control, (2) active control, (3) fixed-dose response, (4) no treatment/concurrent control, and (5) historical control. These are all appropriate in certain settings. Of those types of trials, the one currently most appropriate for traumatic brain injury would be the placebo-controlled trial, because there is no drug approved and recognized to be effective.

The FDA Modernization Act. In 1997 the law was fundamentally changed and is now the FDA Modernization Act (FDAMA). This statute permits many things that the previous law did not, but it does not obligate the FDA to utilize any of these new provisions. The new law states that the Secretary may determine, based on relevant science, that data from *one* adequate and well-controlled investigation, and confirmatory evidence, are sufficient to establish effectiveness. The Secretary may consider such evidence as "substantial evidence."

However, there are some important issues relating to the new law that must be pointed out. We have no idea from the statute, or from the regulatory history, when it would be acceptable to establish a ruling of "substantial evidence" of effectiveness on the basis of a single trial. We do not know what "confirmatory evidence" means. At least, we do not know what Congress had in mind when they passed the law. The agency has developed a document that addresses the circumstances under which one trial might be sufficient to establish effectiveness. It would have to be a large trial that establishes an effect on mortality or very serious morbidity. What would be the characteristics of such a trial? There would have to be an internal replication in a single multicenter trial: many centers showing significance of treatment, perhaps several of them showing statistical significance on their own. The trial might show that there were multiple subgroups of patients within the study that had a similar result. There would be very low *p* values, suggesting that the trial was not positive by chance. Perhaps there would be

multiple outcomes, all showing statistical significance. Also, one might have a trial that could not or should not be repeated on ethical grounds.

Our view is that this provision, while it can be used, is rightly seen as a deviation from the standard. Only in those cases where we really cannot repeat a study would the "one trial" approach be appropriate. We ordinarily require that there be at least two trials showing an effect on appropriate outcome measures. This requirement for replication is to make sure that the drug effect is not due to chance, bias, or perhaps even fraud, and that the effect is real.

Crossover indications. A request is sometimes made to approve a drug for a new indication, based on studies for a different indication. Generally speaking, we do not find that terribly compelling; however, there are situations where we have allowed one trial to suffice for substantial evidence of effectiveness. It has been done with anticonvulsants. For instance, if a drug has been shown in two adequate and well-controlled trials to be effective as an adjunctive therapy, and the company wants to have it approved for monotherapy, it is possible that a single trial on monotherapy would be sufficient. For a certain seizure type, if there are studies showing an effective therapy in adults, a single robust trial in children could be adequate to support that indication. The FDA looks at the totality of evidence in making a decision on what is necessary. If you are looking at two entirely different diseases, for example, stroke and head trauma, I think those disorders would require separate trials. One trial in TBI and one in stroke would not constitute substantial evidence.

Fast track approval. Another new provision that is applicable to head trauma is called "fast track". The pharmaceutical industry certainly knows about this: since November 1997 every drug that has come to the FDA has been proposed for the fast track. FDAMA says that the Secretary shall facilitate the development and expedite the review if the drug is intended for the treatment of a serious or life-threatening condition, and there is the potential to address unmet medical needs. A drug for TBI is a perfect candidate for being "fast tracked." It obviously is intended for the treatment of a serious or life-threatening illness, and would fulfill an unmet medical need. While this is a new provision in the law, there have been regulations concerning these sorts of drugs for a long time. Some of you might have heard about "subpart E," which was an attempt in the regulations to expedite the development and approval of drugs intended to treat such illnesses if the drugs had an effect on mortality or irreversible morbidity. We have routinely classified (at least provisionally) drugs to treat stroke and head trauma as subpart E drugs. The value of being a subpart E drug is to encourage a development program whereby if the adequate and well-controlled trials were positive, one would not have to do extensive Phase III safety testing. If the drug had a beneficial effect on "extreme outcomes," that is, irreversible morbidity and mortality, we would not need to wait for another couple of years to see the complete adverse event profile. We would be willing to approve something that had a big effect relatively quickly.

It is important to realize that nothing in the fast track law or regulations should be interpreted to mean that the usual requirements for substantial evidence are suspended. I bring this up because often sponsors will come in and say, "Well, we have a fast track drug (or subpart E drug), so we only need one study." That is not true. It *may* be true under the new provisions of FDAMA—one study and confirmatory evidence—but the default position is two studies, even for a fast track drug or a subpart E drug.

A great advantage of being designated fast track is that the FDA must review the new drug application within six months of submission, whereas the standard review time is ten months. That only says we would have to take an action, it does not mean drug approval. It might be a provisional, "approve-able" action that would require the sponsor to do more work.

Surrogate endpoints. The FDAMA also introduced into law the concept of potential approval for a drug based on its effect on a surrogate endpoint. The regulations have talked about surrogate endpoints since about 1992, but now it is in the law. It says that drug approval can occur upon a determination that the product has an effect on a surrogate endpoint that is reasonably likely to predict clinical benefit. For example, antihypertensives or cholesterol-lowering agents are drugs for which trials have shown an effect on a surrogate endpoint. The endpoint in and of itself has no clinical meaning; no one is symptomatic if their blood pressure is slightly up or their cholesterol slightly elevated. The reason that the agency has approved such drugs is that there has been overwhelming evidence, based on long-term outcomes, that lowering blood pressure or cholesterol is good.

But the law says "reasonably likely to predict clinical benefit," which is a very different phrase. It says we do not know if the surrogate is validated, but the Agency has the authority to approve drugs on the basis of surrogate outcomes. At the time this was written, people at the FDA were very nervous, because they thought this provision violates the requirement for substantial evidence. The Agency interpretation is that you still have to have substantial evidence of efficacy; it just has to be for the surrogate endpoint. Nevertheless, it still does not permit the use of surrogates that have not been validated against long-term outcome.

Why are surrogates useful? If the outcome you really care about happens well into the future, for example years from now, you cannot practically do those studies. In such instances, you try and identify a marker that can be quantified early, and then demonstrate that it predicts the outcome. The regulation also says that when the clinical effects are easily measured, nonvalidated surrogates are not likely to be acceptable. It gives as examples epilepsy, depression, and psychosis. Trials for these indications are certainly do-able, and we approve drugs on the basis of their effects on the clinical measures. In traumatic brain injury the effects on clinical outcomes are also measurable fairly quickly. These are not studies that need to be years long. If you consider outcomes at 3-6 months after injury, those studies can certainly be done. Hence, surrogate endpoints are harder to sell as the primary outcome measures in TBI trials.

What are the problems of surrogates and why are they treacherous? The treatment may have an effect on the surrogate, but may be dissociated from the disease-causing mechanisms. If there are many different mechanisms, as seems to be the case in TBI, the surrogate may be in one pathway and what you really care about is in a different pathway. Tom Fleming, a statistician at the University of Washington, has written about the problems in validating surrogates. Recently at a meeting he gave a rather extreme example to illustrate the point. People who smoke have discoloration of their fingernails and they die of lung cancer. If you put gloves on a smoker you will have an absolute effect in eradicating their fingernail problem, but you are not going to do anything for the outcome that matters. The big problem in aiming trials at surrogates is that the treatment may affect the clinical outcome in unintended negative ways, but may have an expected, beneficial effect on the surrogate. An example of this would be a trial that looked at the effect of a drug on suppressing arrhythmias, as a surrogate marker of efficacy in cardiac patients. The drug did in fact suppress arrhythmias, but there was a higher mortality among those patients than among the controls. Had we relied on the effect on the surrogate, we would have been very wrong in approving the drug. The consideration is whether or not the effect of the treatment on the surrogate uniformly predicts the effect on the outcome. It has to be much more than a correlation.

Another problem with validating surrogates is more complex. It might be the case that the drug has the intended effect on the surrogate and for the outcome that you really care about, but the effect may be drug-specific or specific to a pharmacological class of drugs. In order for a surrogate really to be validated there should be a robust finding across all relevant

pharmacological classes of drugs showing that the surrogate and the outcome of interest always go in the same direction.

The take-home message is that surrogate outcomes are very difficult to validate. There are, however, some factors that might make a surrogate acceptable, and it has to do with the biological plausibility. In other words, there is a lot of evidence: animal models robustly show the effect, there is a well understood disease mechanism, the mechanism of the drug is well understood, and the surrogate occurs late in the pathway of the pathophysiology. The agency will also consider the public health when looking at approval based on surrogates. Such considerations include: the treatment is directed at a serious illness, there are no treatments available, it is very difficult to study outcome endpoints, and there is a large safety database. Those are compelling reasons to rely on the surrogate, and that is, I believe, the situation that we have for TBI.

International versus U.S. trials. It is perfectly permissible for the Agency to approve a treatment entirely on the basis of non-domestic studies, but there are a number of caveats. The population studied would have to be similar to that in the U.S. population, and the sponsor would have to make that case. The standard of care would have to be very similar to the United States. There may be drugs used routinely in the foreign studies that are not even available here. The standards of record keeping would have to meet U.S. requirements and be available for inspection. It is an integral part of our review to actually go and look at the primary records and ensure that things are as the sponsor says they are. There is a new document out by the International Committee on Harmonization, (ICH) which says that the host country, in this case the United States, can ask for a trial in its country for no particular reason. We would certainly ask for a domestic trial if the foreign data were from a place where we had no experience. In one recent example a country had undergone a revolution during the study, and the documents were unavailable. That struck us as rather problematic. However, we do see many multinational trials that include US and foreign sites, and it is usually not a problem.

Question: You mentioned the concept of not being able to repeat a trial. Would you say more about that, considering a situation in which the medical profession is happy with one trial and considers doing a second placebo-controlled trial unethical?

Answer: There have been cases where we have not been happy with trials that the academic community liked, and that can be a problem. But, generally speaking for large trials with a robust effect on mortality or another definitive outcome, I don't think the agency would try to enforce a higher standard than the scientific community accepts. We cannot say that you have to do another trial and look at mortality if no one is willing to do such a trial. I have just recently come from a meeting with another community of experts, who want active-control trials. They have some drugs that are approved for their indication and they want to show equivalence of the new and the old. Generally speaking, we think those studies are not interpretable, and cannot be used to support a labeling claim.

Question: In the happy circumstance that one gets a positive result in a particular trial, and there are other ongoing trials, what is your philosophy about whether those ongoing trials would have to change in order to reflect the newly found knowledge in the area?

Answer: It may become inappropriate to study a new drug against a placebo control when a drug is approved. However, the scientific community may believe that even though the drug has been approved, that its effect is minimal and they are therefore unwilling to switch their patients to that drug. It is the case in Alzheimer's disease. It has recently been true in MS, although this may change. When a drug really has only a symptomatic effect, has no effect on the underlying disease, or does not prevent any real serious harm, I think people feel very comfortable in doing placebo control trials even though some drugs have been approved.

Generally speaking, we would ask the sponsors to change the informed consent in other ongoing trials to tell patients what their other options are.

Question: How do you go about determining that the community is not willing to run another placebo-controlled trial?

Answer: We might initiate it, if the effect were truly spectacular. Most of the drugs we approve do not have those huge effects, and it has not really been a problem. Another way we would know about it is through meetings like this. Sponsors tell us. Sometimes the study designs that sponsors propose, because they believe placebos are unethical, are trial designs that will give uninterpretable results.

Question: Is there any particular advantage to a disease having an orphan status?

Answer: Advantages accrue to a sponsor whose drug applies to an orphan disease, that is, prevalence of less than 200,000 cases. I believe they get seven years of exclusivity. They are exempted from the requirement to do pediatric studies. They are not exempt from the standards of law for approval, which is substantial evidence of effectiveness. They still have to do the trials. We would certainly negotiate with a company about how much data they would need, given the actual prevalence of the disorder.

REPORTS FROM WORKING GROUPS

For the second day of the workshop, the participants were divided into two large working groups to discuss and make recommendations on four areas of particular concern in the conduct of TBI trials. These areas were selected by the organizers ahead of time and assignment of membership made at the end of the first day. The chosen topics were preclinical testing, clinical trial design, outcome measures, and surrogate endpoints. Reports from the working groups are presented below, as summarized by the discussion leader.

Preclinical Testing (Tracy McIntosh, Ph.D., et al.)

This subgroup discussed the best ways to test drugs in animal models to ensure a thorough evaluation before recommending them for clinical trial. Our conclusions are summarized below:

Targeting mechanisms. The group felt that more research is needed to identify pathophysiological mechanisms of TBI and then to generate compounds that are specifically targeted to these mechanisms. The new fields of genomics and proteomics are greatly enhancing our ability to understand these basic mechanisms. We need to determine whether these mechanisms are common between animal models, between models of diffuse and focal injury, in subarachnoid hemorrhage and in contusions. The group agreed that it would be desirable to target therapies towards these different types of injuries.

Extent of preclinical testing. How extensively should a drug be tested in preclinical models prior to the initiation of clinical trials in TBI? The group felt that a compound should be evaluated not just in one model, but also in multiple animal models. It is preferable that two or more different labs, ideally using different experimental TBI models, evaluate compounds. At least one of the labs should be independent from the company that is developing the drug.

Animal models. There are excellent clinically relevant animal models that have been established in rodents, for example, the lateral fluid percussion model. The group had long discussions about the utility of moving up to more gyroencephalic species (larger animals) to test the compounds that seemed promising. There was a consensus in the group that if

something is likely to work clinically, it needs to have a robust effect in several different animal models.

It was felt that the pig is a good model to move to after the rodent, because it reproduces human anatomy fairly well. However, it was recognized that moving to higher species might be problematic, especially for investigators working outside the United States. While recognizing and acknowledging this problem, it was nevertheless felt that there is a need to develop and characterize trauma models in these animals (e.g., pigs or primates). The model needs to imitate the mechanism of injury, should be reproducible, and include behavioral endpoints. The development of adequate, appropriate, and validated behavioral indices in these large animal species will be a major challenge to the field.

Window of treatment. The third point for discussion was the concept of a "critical window of opportunity" during which a drug may be expected to have a positive effect. There is truly a lack of knowledge of the critical window, both in humans and in animals. How the critical window in an animal model correlates with the window in humans in largely unknown and may well be different for each drug.

Research priorities. Based on discussions between the basic scientists and clinicians in the group, the following research priorities were articulated:

1. Model traumatic coma. Most of the currently used models represent mild to moderate TBI.

2. Establish animal ICUs to be able to study severe TBI and to look at surrogate markers of outcome.

3. Study pharmacokinetics in multiple animal models in order to establish how much of the drug needs to be given, how soon, how often and for how long.

4. Evaluate drug interactions, particularly considering the variety of treatments given to head injury patients.

5. Evaluate drug cocktails aimed at the different mechanism of injury.

6. Study age-related differences, including pediatric and geriatric TBI.

7. Focus more study on late (chronic) events in animal models.

8. Model field or ER therapies, targeting clinically relevant secondary insults, such as hypoxia, ischemia, hemorrhage and hyperventilation.

9. Evaluate female/male differences, not only with respect to drug efficacy, but also in terms of outcome differences without pharmacological therapy.

Minimum preclinical development. One of the main charges to the group was to develop a set of criteria that were the minimum needed before a compound could be deemed ready for clinical trials. The following criteria were agreed upon:

1. Establish efficacy in multiple rodent models of TBI. Although it is desirable a to also show efficacy in larger animals, the consensus was that this may not always be practical.

2. Establish the pharmacokinetics of the drug and use this information in establishing the clinical protocol.

3. Evaluate the long-term effects of the drug both physiologically and behaviorally in the animal models.

4. Establish efficacy in rodent models ranging in severity from mild to severe.

5. Utilize more sophisticated statistical design and analysis in preclinical trials.

6. Establish greater communication between the basic scientists, the clinicians, and the pharmaceutical industry in reaching decisions regarding drug development.

Question: Are you suggesting that if you showed a replicated effect in a model of predominantly white matter injury, but not in a model of gray matter injury, that you might not go forward with the agent?

Answer: Our group would likely suggest that it might be more useful to test that compound in patients who show predominant white matter injury.

Question: The human head is very mobile versus the trunk, and is very heavy (especially in children). This is likely a cause for the inertial loading injury, and I do not see any animal model that is quite appropriate. I wonder what species comes close enough to look at this kind of injury?

Answer: There is a need to incorporate bioengineering into our modeling, and there are bioengineers interested in this field. We need to cultivate that interest and bring them in to our laboratories to address such issues more thoroughly.

Question: What kinds of increased statistical involvement in design of preclinical trials does your committee have in mind?

Answer: Many of the preclinical studies to date have been performed using a simple, one-way analysis variance model. Many times it is very interesting to look at a combination of factors (for instance, an experimental agent tested at two different severities of injury). Rather than creating four or five different studies, we could do a two-way design, which would allow examination of the interactions.

Question: I am surprised that the optic nerve stretch model has not had a bigger place, because it produces axonal injury in a biomechanically defined way. Is it not a model where much greater progress can be made, rather than trying to model diffuse axonal injury in the "whole head" of small animals?

Answer: The optic nerve stretch model is a relatively pure model of axonal damage, and it is an *in vivo* model, so one can evaluate pharmaceutical compounds. However, our group was focused on recommendations for the preclinical studies that would provide the most comprehensive and relevant data with respect to future clinical studies. For that we need not only mechanistic data, but also behavioral endpoints. The optic nerve stretch model unfortunately does not provide these.

Question: You mentioned evaluation of concurrent medication to reproduce the clinical setting. What about initial sedation, anesthesia, or paralysis?

Answer: Most labs do not standardize their anesthesia paradigms, and different anesthetics and sedatives can have profound effects on the way animals respond to the traumatic injury. More attention needs to be paid to this.

Comment: It would be very interesting and important to look at previous drugs that failed in clinical trials and to test them in animals in the way that they were used in the hospital. There is a big difference between how compounds are tested in animals and in humans. If we gave a rat four doses a day of Selfotel for a week, what would be the results?

Answer: We have recommended the incorporation of secondary insults such as hypotension and hypoxia into the models, and looking at drug effects in a more clinically relevant context.

Clinical Trial Design (M. Ross Bullock, M.D., Ph.D., et al.)

The ideal design of clinical trials in TBI remains to be defined. However, much has been learned from the trials conducted to date. After extensive discussions, the group endorsed these general principles for future clinical trials:

1. **Demonstrated mechanism:** A mechanism of injury should be demonstrated in animal models as well as in human TBI. The test drug should be active against that mechanism.

2. **Brain penetration:** The drug should be shown to penetrate the brain in appropriate concentrations.

3. **Safety profile:** The drug should have been demonstrated to be safe in human TBI. In addition, the drugs should be shown safe in combination with other agents that are currently used clinically, such as anticonvulsants.

4. **Time window:** The drugs should be administered in the appropriate time window. We have to define that time-window not only in rodents, but also in the gyrencephalic animals and in humans. Treatment windows are a major issue in the conduct of these trials. Early treatment seems desirable. We need to move towards conduct of trials in the prehospital setting, with EMS or ER personnel giving the trial drugs. As we do that, we will come up against the issue of consent. Most of the time, it will not be possible to get full informed consent in these circumstances. We should revisit the "waiver of consent" issue. Clearly, the government agencies are concerned with this area, and we need to consult with them to define approved ways to use waiver of consent for head trauma patients. A combined meeting between representatives of the funding and oversight agencies with ABIC could be valuable. Other interested groups would include those who study epilepsy, cardiac arrest and emergency medicine and bioethics. Organizations such as the Brain Trauma Foundation and the Brain Injury Association can also play an important role.

5. **Alternative trail designs:** Alternatives to the conventional prospective, randomized, controlled trial were considered. Mega-trials are very interesting, but only need to be considered when looking for small drug effect. Most of the trials that have been conducted in TBI have been over-ambitious in setting the expected treatment effect at 10% improvement. This expectation has been based on animal studies, which have sometimes yielded dramatic separation between control and drug groups. It has been subsequently assumed that our error was in underpowering the trials (too few patients to show the desired effect). The actual error may well have been in expecting an unrealistically great effect! A serious issue for mega-trials is that there are no appropriate assessment tools for moderate and mild-injury. It is impractical to use complicated neuropsychological testing for 10,000 subjects, and the GOS is probably inappropriate for this patient group since 95% of subjects will probably be found to have a "good recovery." More thought needs to be directed at outcomes issues in moderate and mild TBI. We have been saying this for three successive meetings on outcomes, so we need to do something!

Another alternative to the conventional prospective, randomized clinical trial (PRCT) is subpopulation-focused trials, for example, the Bradycor trial. These trials have got to be based on adequate and well-analyzed phase II studies to identify the mechanism of interest. What would be the optimal trial design for a PRC? No generic boilerplate design can be proposed, and we should tailor the protocol specifically to mechanisms that emerge from animal studies,

and are tested in phase II studies. There are adequate strategies available to make sure that the treatment groups are equally randomized.

6. **Careful monitoring of management:** There needs to be aggressive and proactive feedback to individual centers on protocol violations, data inadequacies, management deviations and so forth. That will require a more dynamic role for the PI and the sponsoring agency conducting the trials.

7. **Stopping rules:** The group felt that trials should be completed if at all possible. The sponsors need to guarantee the investigators that they intend to take the trials to completion. Interim analyses are appropriate in guiding the conduct of the trial and safety of patients. However, over use of futility analyses may end a trial prematurely. These issues need to be clearly debated with sponsors before the trial starts and there must be predetermined stopping rules. Sponsors should agree to independent safety monitoring boards, to oversee the conduct of the trial. This is a necessary for the safety of the patients and to maintain scientific credibility.

8. **Publication of results:** Independent and full analysis of the data should be the standard. The guidelines for analysis and publication should be clear, and papers should be prepared for publication within a year of the completion of the trial.

Comment: There are outcome measures available for moderate and mild head injury. It may be possible to evaluate at 1-3 months, use the GOSE, and not dichotomize but use the full range of outcomes. We (Temkin et al.) have a new measure that will go down to the lowest level of impairment, and compare patients to themselves not the general population, in any of nine areas. We have used this assessment in the valproate trial, which was not developed to look for a neuroprotective effect. This measure and the GOSE are being used in our current magnesium sulfate study.

Question: Could you talk for a minute about what the specific problems will be with the exception from informed consent procedure?

Answer: These patients arrive at the hospital in a highly emergent situation: There are many things being done to them simultaneously, and it is often very difficult to superimpose informed consent upon all this. Moreover, the severe TBI patient is by definition comatose, and very often there is no family member immediately available to give consent. If it is shown that a drug needs to be given within 2 hrs of the injury, either the emergency squad has to administer it at the scene, or it has to be given soon after the patient enters the ER. Most of the time, there is no family member at the scene of the accident or waiting in the ER to give consent. Hence, if research in this field is to move forward, a mechanism for waived consent needs to be established.

Question: Could you comment on waiver of consent rules? Are they complicated?

Answer: Since these new regulations have been published, two drug trials have been started in which the pharmaceutical companies chose not to request waiver of consent. The OPRR (Office of Human Research Protections, DHHS; this office no longer exists. The current office most relevant to these issues is the Office for Human Research Protections, DHHS.) created such complex hurdles that the companies decided that this was not a battle they wanted to fight. These regulations make it so unpalatable that the "waiver option" just is not being used.

Question: I am interested because we in the FDA never get to hear about this. What specifically are the major hurdles? Is it the community consultation or locating family?

Answer: The interpretation of "community" is the issue! How do we know that notice been given to the community? How many talk shows do you have to go on and how many churches

do you need to visit? Is that really the population you want to reach? All of these issues remain undefined and could be challenged. If you have waiver of consent available in a multi-centered study, and one IRB refuses to accept it, that decision must be promulgated by the pharmaceutical company to all of the other IRBs concerned. It creates a level of uncertainty about completion of the trial and center participation. The justification is unknown for these regulations.

Comment: Our center (San Diego) entered the third largest sample of patients in the Tirilazad trial in the United States, but we are now "out of business" for severe head injury trials, because of the rules of waiver. We have not been able to satisfy these regulations in a meaningful way. They are burdensome. They obstruct research. I chaired the IRB in San Diego for five years and I can tell you that this is the single biggest impediment to finding new treatments for TBI and other acute conditions. It is really unethical if you cannot figure out a way around this to continue to do research in cardiac diseases, for example. Somebody with acute chest pain— can he/she really give informed consent?

Answer: Researchers are concerned that these analyses are more often used to provide support for an economic argument to stop a trial, and not share the data with the scientific community. We are not learning anything from such studies to help us in the future. I agree it is unethical to put patients at risk for a drug that is harmful. When there is a high probability that you will not find the difference you were powered for, surely you should not put them at risk for getting the drug at all? We are only talking about that point where the null hypothesis is liable not to be rejected.

Comment: I think using a futility analysis in a circumstance where there are no treatments for a disease, loses information. So is the fact that a trial will not show a 10% effect, a reason to stop? If you go to completion you may see a difference of 5%. That is potentially important information.

Comment: Futility analyses are tools that should be used by DSMBs (Data Safety Monitoring Boards) in making appropriate decisions about a research risk for the patient.

Outcome Measures (Claudia Robertson, M.D., et al.)

There was a lively discussion with many opinions expressed, but the group did come to agreement on several points. The consensus can be summarized as follows:

1. In general, for an intervention or a treatment to be accepted into general clinical practice, investigators should show benefit on an index relevant to lifestyle.

2. The size of the benefit and the exact outcome scale should be related to the population and the treatment being studied. These scales can vary, and the utility of the measures in TBI patients requires ongoing investigation.

3. An evaluation at 6 months after injury probably provides the best assessment of the ultimate outcome and is least influenced by the confounding effect of other associated injuries. There is general enthusiasm for using the GOSE.

The rationale behind the last recommendation is that the GOSE provides a better distinction between levels of disability, particularly with the severe disability group, while the GOS tends to lump a wide range of disability into one category.

There was absolutely no enthusiasm for economic assessment (e.g., length of stay or cost of stay) as part of outcome assessment. We heard earlier that often the acute cost of hospitalization could be more expensive in someone who might ultimately have a better outcome. This issue could be very confusing if added in as an important outcome measure.

We noted that outcome measures might have to be designed for the specific goals of the patient populations who will be included in individual trials. Every trial that we do is a little different, and every patient population is a little different, depending on the inclusion/exclusion criteria. We cannot rigidly apply the same outcome measures for each trial. There may be trial circumstances where a dichotomized GOS might be best divided between good recovery and moderate disability, and other situations where the cut might need to be made between moderate disability and severe disability.

There is also considerable enthusiasm for the concept of comparing observed outcome in a trial to outcome predicted by initial injury severity. Teasdale and Maas described this kind of analysis at this meeting, and the group would like to see it implemented in future trials.

In general, there was agreement that we have been overly enthusiastic, or overly optimistic, in designing trials to achieve a 10% improvement in the overall outcome. We should seriously consider using 5-7.5% improvement for upcoming trials.

Question: On your suggestion for using actual outcomes versus predicted outcomes, did you all feel that is something that could be translatable from one center to the other or would you need to develop a prediction model based on data derived from individual centers?

Answer: I suspect that each trial would require some kind of prediction. The inclusion and exclusion criteria would be different for each trial, and there may be unique aspects for a particular trial. But I am not sure that it would have to be specific to each center within a trial. These are things that will have to be worked out. In general, this may be the way that trials should be designed in the future.

Question: Are you saying that predicted outcomes could be used instead of controls?

Answer: No, absolutely not.

Question: I would like to ask Dr. Katz how the FDA views that. Consider that we have lowered the hypothetical differences between the treatment/control groups to 5%, and the study was large enough to show statistical differences. What is the acceptable difference between two groups to get a drug approved?

Answer (Dr. Katz): There are no *a priori* levels. If the medical community felt that this sort of a difference was something that mattered clinically, we could be convinced. Any final decision about approving a drug will be a risk/benefit comparison. We look at the data closely to see what the risks and the benefits really are.

Comment: Dr. Katz made an important point: where you set the bar is really clinical opinion. I think that the 10% bar was established rather arbitrarily based on hope. I think that we may be able to achieve a more targeted outcome with fewer patients—that may support the 10% difference.

Comment: Whatever is done for the first approved treatment will set a precedent. The result will become a standard in the field that will be very difficult to change. I think we need to give considerable thought to what that standard ought to be. Sometimes, if there is nothing available, you set the standard relatively low, just to get something out there. But this approach perpetuates itself. Several "me-too" drugs come along using the same standard of marginal efficacy, and the field becomes full of treatments of marginal significance. I would suggest that we want to give some serious thought to whether or not we embark on looking at smaller effects and whether that is meaningful for patients.

Surrogate Endpoints (Graham Teasdale, M.D., et al.)

The second task of this group was to look at surrogate endpoints, and our discussion yielded diversity and consensus. It was agreed that outcomes or other surrogate measures observed in the first two weeks after injury could not provide a satisfactory index of benefit, and support the clinical adoption of a treatment. Nevertheless, early indicators can be of interest in two ways:

First, they may provide confirmation that treatment is in some way affecting the biology of the injured brain; this information is useful as a "proof of concept." Second, surrogate measures may be useful in assessing dose-ranging effects by providing a biological marker that indicates drug activity or adverse effects. For example, while using a drug that is aimed at reducing the ICP, one can decide upon the appropriate dose range based on the effect on ICP and blood pressure.

When using some early information to predict later benefit it is required that the surrogate predicts the eventual outcome, not just that it shows a general relation to outcome. The other concept that came through was that any correlation between an early effect and change in later outcome would need to be revalidated, perhaps using more than one index, in more than one study. It would also be essential to repeat this process of validation before coming to a conclusion about another new agent. For example, it would not be sensible to translate an early ICP effect from drug X to a similar benefit of drug Y, if X affects cerebral blood volume and Y affects axonal damage.

One value of an early surrogate outcome could be in shortening the time to obtain an answer, either in a single patient or in a trial. Nevertheless, this reduction in time frame is probably not so important in head injury. A shorter time is essential if the eventual answer takes 5-10 years to obtain; for example, survival in cancer or AIDS. Shortening the observation period in a TBI trial from 6 months to 1 week may not be a sufficient advantage. A surrogate may also be useful in obtaining an answer in fewer patients if there is truly a large effect. Such information might be valuable in a sequential trial design, providing an early index from which to decide whether to continue the trial, or to change the design.

Will a surrogate index be sufficient in and of itself as evidence in support of final drug approval? If a surrogate is to be used, what should its features be? The index should produce data that are quantifiable in a well-calibrated way so that it indicated the severity, extent, and type of brain damage. Thus, measures that only reflect secondary effects of brain trauma, such as ICP or CBF, have their limitations. It was also agreed that the measure should be dynamic, so that it is possible to observe changes in the individual patient in response to intervention, preferably in a way that is related to the mechanism of the drug.

With further discussions it was recognized that there are other important technical features of an early index: how easy it is to measure, its feasibility in severe injuries, its cost, how frequently it is necessary to make observations, and the complexity of data analysis (e.g., functional imaging). Some candidate surrogate markers and their attributes are listed in Table 5.

Intracranial pressure. It is certainly quantifiable and measurable. It can be summarized in many ways (e.g., percent time at different levels; mean value). Unfortunately, it has only an indirect relationship to the severity of brain damage, and primarily reflects the amount of brain swelling produced by the trauma. Does it relate to the severity of axonal injury? Axonal injury may be the main determinant of the quality of recovery in survivors. Even though high ICP is a good acute predictor of mortality, it may not be a good predictor of the quality of recovery in survivors. ICP is dynamic: one can see changes, and they depend upon the treatment. In

general, if the drug is targeted to reduce ischemic brain damage with its consequent swelling, but does not show an ICP effect, it is unlikely that the drug will work. ICP measurements certainly have to be considered to be useful, but may not an entirely reliable index if used as a sole indicator.

Therapy intensity level. We recognized that there is some interaction between ICP and the type of treatment a patient receives; they should be looked at together. It is possible to rank therapy intensity level (TIL), but that is not quite the same as quantifying it is a well-calibrated way. Any relationship of TIL to severity of brain damage is indirect, and it is too "operator-dependent" to provide consistent information. It is probably necessary to record TIL to interpret ICP data, but by itself TIL is not much use as a surrogate.

Jugular venous oxygen saturation. It is possible to count how often episodes of jugular venous oxygen desaturation occur, how long they last, and how severe they are, so that quantification is achieved. However, it is not absolutely clear how they relate to the quantum of brain damage. Do desaturations indicate the threat of worsening brain damage, or reflect how badly the brain is already damaged? They are dynamic, but is it possible to say that they have been prevented in an individual patient? If the intervention is designed to raise CBF, its effect should be reflected in "higher" $SjvO_2$, but this will be the case for an agent working on various cellular mechanisms.

Structural imaging. It is possible to quantify the number of contusions, the extent of midline shift, whether or not the basal systems are compressed, and so on. As a consequence, this approach can be useful for space-occupying lesions. In contrast, in even fatal diffuse axonal injury, imaging can be completely normal. Imaging does not lend itself to dynamic measurements. Lesions or brain shifts can get bigger or smaller, but these changes take place over long periods and may have very little relationship to the mechanisms of a pharmacological agent. It is difficult to conduct imaging studies in acute, severe head injuries. Methods such as SPECT or PET may be very useful as research tools, but they are not easy to use in critically ill patients.

Metabolic measurements. There are a lot of data about metabolic measures, but how these should be interpreted is less clear. The results may reflect severity of injury in some brain areas, but any relationship to overall damage remains to be shown. Measurements can be dynamic, but they may show irrelevant changes in metabolites or drug concentrations.

Neurological worsening/neurological improvement. Worsening either happens or it does not, so it may or may not be quantifiable. Improvement (time to reach a certain index or proportion of patients who are at a certain level at a particular time) might provide graded information. These events are dynamic, and they are likely to relate to the amount of brain damage. They will describe how long the damage lasted or how quickly it resolved, but the measures are probably not related to a drug mechanism.

Summary. Our group felt that no single "surrogate" was ideal, and though useful information can be gained, there is need further development. Neurological worsening/improvement might be components of a quality assurance system for different centers or might be early indicators of improvement and recovery.

Question: Did the group discuss use of these outcomes in phase II trials? If an effect on a certain parameter is seen in phase II, should that measurement be a part of the phase III trial?

Answer: These surrogates by themselves were considered not to be primary endpoints for phase III. They were seen as quite valuable for phase II. We did not discuss whether or not they were

necessary and essential in phase III. Their place in phase III could be in a sequential design, to produce some early signal whether to continue.

Comment: Imaging is probably not dynamic, but if its use could shorten the time period of assessment from several months to a week, it would be valuable.

Answer: Yes, but you can only image the people who are still alive.

Comment: I think that you are "downgrading" the dynamic aspect of imaging because you are only thinking in terms of acute injury. What about the brain picture at 2 weeks or 1 month, versus the 6 months necessary for behavioral assessment? Perhaps using spectroscopy or magnetic transfer imaging may predict outcome.

Answer: At present, MRI is expensive, cumbersome, and difficult to perform in TBI patients, even after 1 wk; and the data are difficult to interpret. We did not think imaging would be crucial in the immediate future. There are opportunities for further improvements in techniques.

Question: What about MR spectroscopy or PET?

Comment: At the present, I think that your comments on the difficulties of imaging the head injured patient in the acute phase are correct; but the results of research are very promising. On the basis of our (Marmarou) spectroscopy work, there is a strong prognostic value for release of *N*-acetylaspartate and recovery. In patients who do not do well, values drop acutely and stay low; in patients who recover, the values increase. Perhaps such an index assessed within a 10-day period would give us a handle on what will occur 6 months later.

Comment: Imaging techniques are fertile areas for research, and could move the field forward through both basic and clinical studies. Imaging cannot be the primary endpoint in a Phase III trial, but the techniques hold promise for elucidating mechanisms that would be the targets for drug trial design.

Answer: There are a number of biochemical or metabolic markers that seem to relate quite well to the amount of damage, but whether those are sufficiently dynamic is a question. It has been suggested that late rises in S100 protein concentrations in the blood might correlate with secondary insults.

REFERENCES

ALDERSON P, ROBERTS I. Corticosteroids in acute traumatic brain injury: systematic review of randomized controlled trials. BMJ 1997;314:1855–1859. [PubMed: 9224126]

BIEGON A, BAR JOSEPH A. Development of HU-211 as a neuroprotectant for ischemic brain damage. Neurol. Res 1995;17:275–280. [PubMed: 7477742]

BRAAKMAN R, SCHOUTEN HJ, BLAAUW-VAN DISHOECK M, et al. Megadose steroids in severe head injury. J. Neurosurg 1983;58:326–330. [PubMed: 6338164]

CHOI SC, MARMAROU A, BULLOCK R, et al. Primary endpoints in phase III clinical trials of severe head trauma: DRS versus GOS. J. Neurotrauma 1998;15:771–776. [PubMed: 9814633]

CHOI SC, BULLOCK R. Design and statistical issues in multicenter trials of severe head injury. Neurol. Res 2001;23:190–192. [PubMed: 11320598]

CLIFTON GL, KREUTZER JS, CHOI SC, et al. Relationship between Glasgow Outcome Scale and neuropsychological measures after brain injury. Neurosurgery 1993;33:34–38. [PubMed: 8355845]

CLIFTON GL, MILLER ER, CHOI SC, et al. Lack of effect of hypothermia in acute brain injury. N. Engl. J. Med 2001;344:556–563. [PubMed: 11207351]

COOPER PR, MOODY S, CLARK WK, et al. Dexamethasone and severe head injury. A prospective double-blind study. J. Neurosurg 1979;51:307–316.

DACEY RG, ALVES WM, RIMEZ RW, et al. Neurosurgical complication after apparently minor head injury: assessment of risk in a series of 610 patients. J. Neurosurg 1986;65:203–210. [PubMed: 3723178]

DEARDEN NM, GIBSON JS, McDOWALL DG, et al. Effect of high-dose dexamethasone on outcome from severe head injury. J. Neurosurg 1986;64:81–88. [PubMed: 3510286]

DICKINSON K, BUNN F, WENTZ R, et al. Size and quality of randomized controlled trials in head injury: a review of published studies. BMJ 2000;320:1308–1311. [PubMed: 10807622]

DOPPENBERG EMR, BULLOCK R. Clinical neuro-protection trials in severe traumatic brain injury: lessons from previous studies. J. Neurotrauma 1997;14:71–80. [PubMed: 9069438]

FAUPEL, G., et al. Double-blind study on the effects of steroids on severe closed head injury. In: Pappius, HM.; Feindel, W., editors. Dynamics of Brain Edema. Springer-Verlag; New York: 1976. p. 337-343.

FEINSTEIN AR, SOSIN DM, WELLS CK. The Will Rogers phenomenon. Stage migration and new diagnostic techniques as a source of misleading statistics for survival in cancer. N. Engl. J. Med 1985;312:1604–1608. [PubMed: 4000199]

GAAB MR, TROST HA, AKANTARA A, et al. Ultrahigh dexamethasone in acute brain injury. Results from a prospective randomized double-blind multicenter trial (GUDHIS). Zentralbl. Neurochir 1994;55:135–143. [PubMed: 7810251]

GHAJAR J. Traumatic brain injury review. Lancet 2000;356:923–929. [PubMed: 11036909]

GIANNOTTA SL, WEISS MH, APUZZO ML, et al. High-dose glucocorticoids in the management of severe head injury. Neurosurgery 1984;15:497–501. [PubMed: 6333649]

GOBIET, W., et al. Treatment of acute cerebral edema with high dose of dexamethasone. In: Beks, JWK.; Bosch, DA.; Brock, M., editors. Intracranial Pressure III. Springer-Verlag; New York: 1976. p. 231-235.

GROTTA J, CLARK W, COULL B, et al. Safety and tolerability of the glutamate antagonist CGS 19755 (Self otel) in patients with acute ischemic stroke. Results of a phase IIa randomized trial. Stroke 1995;26:602–605. [PubMed: 7709405]

GUDEMAN SK, MILLER JD, BECKER DP. Failure of high-dose steroid therapy to influence intracranial pressure in patients with severe head injury. J. Neurosurg 1979;51:301–306. [PubMed: 469578]

GUIDELINES AUTHORS. Management and prognosis of severe traumatic brain injury. J. Neurotrauma 2000;17:451–627.

HALL, ED.; BRAUGHLER, JM. Free radicals in CNS injury. In: Waxman, S., editor. Molecular and Cellular Approaches to the Treatment of Brain Diseases. Raven Press; New York: 1993. p. 81-105.

INTERNATIONAL STROKE TRIAL COLLABORATIVE GROUP. The international stroke trial (IST): a randomized trial of aspirin, subcutaneous heparin, both or neither among 19,435 patients with acute ischaemic stroke. Lancet 1997;349:1569–1581. [PubMed: 9174558]

KONTOS HA, WEI EP, CHRISTMAN CW, et al. Free oxygen radicals in cerebral vascular responses. Physiologist 1983;26:165–169. [PubMed: 6611683]

KONTOS HA. George E. Brown memorial lecture. Oxygen radicals in cerebral vascular injury. Circ. Res 1985;57:508–516.

LAVIE G, TEICHER A, SHOHAMI E, et al. Dexanabinol (HU-211) improves long-term outcome in rats subjected to transient focal brain ischemia. Am. Soc. Neurosci. Abst 2000;30:771.

MAAS AL. Neuroprotective agents in traumatic brain injury. Expert Opin. Invest. Drugs 2001;10:753–767.

MARION DW, PENROD LE, KELSEY SF, et al. Treatment of traumatic brain injury with moderate hypothermia. N. Engl. J. Med 1997;336:540–546. [PubMed: 9023090]

MARMAROU A, NICHOLS J, BURGESS J, et al. Effects of the bradykinin antagonist Bradycor (deltibant, CP-1027) in severe traumatic brain injury: results of a multi-center, randomized, placebo-controlled trial. American Brain Injury Consortium Study Group. J. Neurotrauma 1999;16:431–444. [PubMed: 10391361]

MARSHALL LF. Head injury: recent past, present, and future. Neurosurgery 2000;47:546–561. [PubMed: 10981741]

MARSHALL LF, MAAS AI, MARSHALL SB, et al. A multicenter trial on the efficacy of using Tirilazad mesylate in cases of head injury. J. Neurosurg 1998;89:519–525. [PubMed: 9761043]

MORRIS GF, BULLOCK R, MARSHALL SB, et al. Failure of the competitive *N*-methyl-d-aspartate antagonist Selfotel (CGS 19755) in the treatment of severe head injury: results of two phase III clinical trials. The Selfotel Investigators. J. Neurosurg 1999;91:737–743. [PubMed: 10541229]

MUIZELAAR JP, MARMAROU A, YOUNG HF, et al. Improving the outcome of severe head injury with the oxygen scavenger polyethylene glycol-conjugated superoxide dismutase: a phase II trial. J. Neurosurg 1993;78:375–382. [PubMed: 8433137]

MURRAY GD, TEASDALE GM, BRAAKMAN R, et al. The European Brain Injury Consortium survey of head injuries. Acta Neurochir. (Wien) 1999;141:223–236. [PubMed: 10214478]

PETO R, BAIGENT C. Trials: the next 50 years. Large scale randomized evidence of moderate benefits. BMJ 1998;317:1170–1171.

PIANTADOSI, S. Clinical Trials: A Methodological Perspective. Wiley; New York: 1997.

POLIDORI MC, FREI B, RORDORF G, et al. Increased levels of plasma cholesteryl-ester hydroperoxides in patients with subarachnoid hemorrhage. Free Radic. Biol. Med 1997;23:762–767. [PubMed: 9296453]

POLIDORI MC, FREI B, CHERUBINI A, et al. Increased plasma levels of lipid hydroperoxides in patients with ischemic stroke. Free Radic. Biol. Med 1998;25:561–567. [PubMed: 9741593]

SAUL TG, DUCKER TB, SALCMAN M, et al. Steroids in severe head injury. A prospective randomized clinical trial. J. Neurosurg 1981;54:596–600.

SCHERBEL U, RAGHUPATHI R, NAKAMURA M, et al. Differential acute and chronic responses of tumor necrosis factor-deficient mice to experimental brain injury. Proc. Natl. Acad. Sci. U.S.A 1999;96:8721–8726. [PubMed: 10411942]

SHOHAMI E, NOVIKOV M, BASS R. Long-term effect of HU-211, a novel non-competitive NMDA antagonist, on motor and memory functions after closed head injury in the rat. Brain Res 1995;674:55–62. [PubMed: 7773695]

SHOHAMI E, GALLILY R, MECHOULAM R, et al. Cytokine production in the brain following closed head injury: Dexanabinol (HU0211) is a novel TNF-alpha inhibitor and an effective neuroprotectant. J. Neuroimmunol 1997;72:169–177. [PubMed: 9042110]

SIMON R. Optimal two-stage designs for phase II clinical trials. Controlled Clin. Trials 1989;10:1–10. [PubMed: 2702835]

STATLER KD, JENKINS LW, DIXON CE, et al. The simple model versus the super model: translating experimental traumatic brain injury research to the bedside. J. Neurotrauma 2001;18:1195–1206. [PubMed: 11721738]

TEASDALE GM, PETTIGREW LE, WILSON JT, et al. Analyzing outcome of treatment of severe head injury: a review and update on advancing the use of the Glasgow Outcome Scale. J. Neurotrauma 1998;15:587–597. [PubMed: 9726258]

WHITEHEAD, J. The Design and Analysis of Sequential Clinical Trials: Revised Second Edition. Wiley; Chichester: 1997.

WHYTE, J. Assessing medical rehabilitation practices: distinctive methodological challenges. In: Fuhrer, MJ., editor. The Promise of Outcomes Research. Brookes; Baltimore: 1997. p. 43-59.

WHYTEJDelisaJRehabilitation of the patient with traumatic brain injury. Rehabilitation Medicine: Principles and Practice199811911239LippincottPhiladelphia(3rd ed)

WILSON JT, PETTIGREW LE, TEASDALE GM. Structured interviews for the Glasgow Outcome Scale and the extended Glasgow Outcome Scale: guidelines for their use. J. Neurotrauma 1998;15:573–585. [PubMed: 9726257]

WILSON JT, PETTIGREW LE, TEASDALE GM. The emotional and cognitive consequences of head injury in relation to the Glasgow Outcome Scale. J. Neurol. Neurosurg. Psychiatr 2000;69:204–209. [PubMed: 10896694]
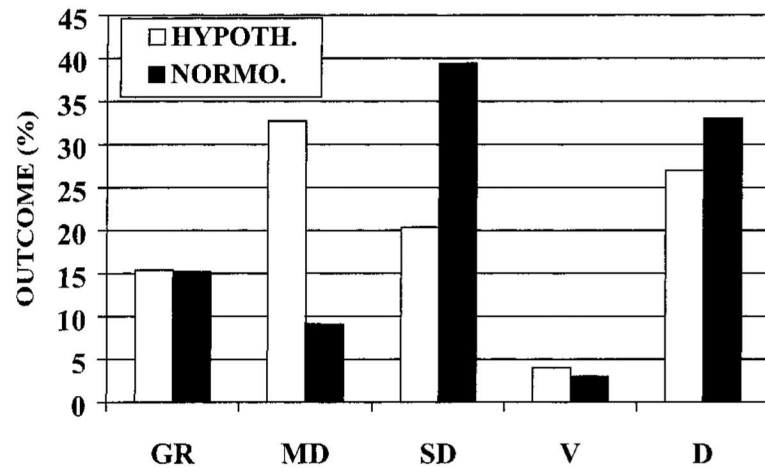
FIG. 1.
Glasgow Outcome Scale scores at 6 months postinjury in patients <46 years of age and admission temperature ≤35°C.

TABLE 1.

Comparison of the Major Differences in the ICP-Targeted (Control) and CBF-Targeted (Experimental) Protocols

	ICP-targeted protocol	CBF-targeted protocol
General measures for all patients		
Sedation	Morphine 5–10 mg q1h prn	Morphine 5–10 mg q1h prn
Controlled ventilation	$paO_2 > 100$, $paCO_2$ 35–40	$paO_2 > 100$, $paCO_2$ 35–40
Normothermia	Antipyretics	Antipyretics
Fluids	Maintenance fluids	Maintain normal CVP(5–8 mm Hg) and normal PWP (8–12 mm Hg)
Treatment goals for physiological variables (all referenced to the same level)		
ICP	<20	<20
MAP	>70	>90
CPP	>50	>70
Management of intracranial hypertension		
CSF drainage	10–20 drops prn	10–20 drops prn
Paralysis	Vecuronium, 10 mg q1h	Vecuronium, 10 mg q1h
Mannitol	25 g q2–6h prn	25 g q2–6h prn
Hyperventilation	$paCO_2$ 25–30	None, $paCO_2$ 35–40
Management of systemic hyper-/hypotension		
Hypertension	For SBP > 160: labetalol	None
Hypotension	For MAP < 70: volume, dopamine, phenylephrine	For MAP < 90: volume, dopamine phenylephrine

ICP, intracranial pressure; CBF, cerebral blood flow; MAP, mean arterial pressure; CPP, cerebral perfusion pressure; SBP, systolic blood pressure; CSF, cerebrospinal fluid; CVP, central venous pressure; PWP, pulmonary wedge pressure.

TABLE 2.

Effect of Misclassifications in Dichotomous Outcomes: An Example

Group	Without misclassifications			With 10% misclassifications		
	Good	Poor	N	Good	Poor	N
Treatment	240	160	400	232	168	400
Control	210	190	400	208	192	400
Difference (%)	60.0	52.5	7.5 [a]	58.0	52.0	6.0 [b]

[a]p = 0.033.

[b]p = 0.088.

TABLE 3.

EFFECT OF HYPOTHERMIA IN SUBGROUP (AGE ≤ 45 YEARS AND ADMISSION TEMPERATURE ≤ 35°C), n = 81

Outcome measures	p*value*
Dichotomous GOS (Favorable, Unfavorable)[a]	0.031
Trichotomous (GR and MD), SD, (V and D)	0.103
Five categories	0.176
DRS	0.196

GR (Good recovery), MD(Moderate disability), S (Severe disability), V (Vegetative), and D (Dead).

[a]Percentage favorable outcomes were hypothermia 52.0% and normothermia 75.8%.

Table 4.

From Bench to Clinic: Where Is the Gap?

	Animal models	Clinical trials	
Parameter	Stroke/TBI	Stroke	TBI
1. Treatment window	Mostly <1 h	3–8 h	4–8 h
2. Treatment duration	Single treatment	Several days	Several days
3. Outcome measures	Infarct size	Clinical (e.g., GOS)	Clinical (e.g., GOS)
4. Follow-up	Mostly 1–3 days	3 months	6–12 months
5. "Adjunct" therapy	None	Multiple	Multiple
6. Study population	a. Young males	Elderly men and women	Mostly young men
	b. Pure strains	Genetic variability	Genetic variability

Table 5.

Surrogate Markers in Traumatic Brain Injury

	Quant	Index of brain damage	Dynamic	Mechanism
ICP	+++	Indirect	√	√/?
TIL	+	Very indirect	?	0/?
SjvO2	++	?	√	0/?
Anatomical imaging	++	+++/0	0	0
MRS	?/++	?/++	?	0
Microdialysis	?	?	√	?
NW/NI	?/++	√	Indirect	0

Desirable features: Quant = degree quantifiable; index of brain damage; dynamic = does the measure change with time or treatment; mechanism = relationship to drug mechanism. Clinical measures: ICP = intracranial pressure; TIL = treatment intensity level; SjV02 = jugular venous oxygen saturation; MRS = magnetic resonance spectroscopy; NW/NI = neural worsening/neural improvement.