Does Methylprednisolone Help Patients with Spinal Cord Injury?

Alex B. Valadka

**Brief Answer**

The National Acute Spinal Cord Injury Study (NASCIS) 2 concluded that a protocol for administration of high-dose methylprednisolone (MP) improved neurologic change scores if it was initiated within 8 hours in patients with spinal cord injury (SCI). These benefits were found not when comparisons were made among all patients randomized to different arms of the study, but only when comparisons were made between patients who received either MP or placebo before the study's median time of drug administration; that is, 8 hours. A comparable analysis in NASCIS 3 led to the recommendation that MP be continued for 48 hours if treatment with MP was not initiated until 3 to 8 hours after injury.

Many criticisms of varying degrees of validity have been directed at the NASCIS trials. Perhaps the most commonly cited concern is that the clinical impact of the improvements in neurologic change scores is minimal. To assess effects on function, NASCIS 3 specifically examined changes in Functional Independence Measure (FIM) scores, but that study was unable to demonstrate significant improvement from administration of MP.

In summary, MP appears to confer some benefit after SCI, but the fact that this benefit has not been shown to improve function makes the use of MP after SCI only a level II recommendation.

**Background**

Since early 1990, administration of high doses of the steroid MP has been promulgated as a standard of care for patients with SCI. The studies that established the efficacy of this treatment, however, have been criticized ever since their publication, and recent years have seen the publication of several additional commentaries that have been quite strong in their criticisms of the original studies. Reviewing the attacks against (and defenses of) the NASCIS trials makes for fascinating reading. This chapter reviews the essential findings of these trials, some of the criticisms that have been directed against them, and the investigators’ responses to these criticisms.

The following sections review the NASCIS trials in some detail. In general, the NASCIS studies were designed to provide class I data, although some of the data analysis and reporting of results may represent a lower class of evidence. Critical commentaries and letters to the editor generally reflect class III data. Readers who are not interested in the details may skip to the “Summary of the NASCIS Trials” that immediately follows the next section.

**Literature Review**

NASCIS 1-2

The original NASCIS trial compared two MP dosing regimens: what was then considered standard therapy
for SCI (100 mg bolus of MP followed by 25 mg every 6 hours for 10 days) versus a regimen of a 1000 mg bolus followed by 250 mg every 6 hours for 10 days. No placebo arm was used because withholding steroids was considered to be unethical at that time. The two groups did not differ in neurologic outcome at 6 weeks, 6 months, or 1 year after injury. Although this is largely considered to be a "negative" trial, it laid the groundwork for the design and analysis of subsequent NASCIS trials.

NASCIS 2: Six-Month Report

As NASCIS 1 progressed, new animal data suggested that higher doses of steroids administered for shorter periods might provide efficacy with fewer complications. These modifications were incorporated into NASCIS 2, a study that used a placebo group and that also included a group that received naloxone, which had proved promising in preclinical studies.

RELEASE OF RESULTS

The 6-month outcome data from NASCIS 2 were published in the New England Journal of Medicine on May 17, 1990. In an unusual move, the Journal agreed to lift its customary embargo concerning prepublication release of a study’s findings, partly because the National Institute of Neurological Disorders and Stroke requested early release in the public interest. The results of NASCIS 2 were widely reported by the media roughly 6 weeks before their publication in the Journal. Thus, a month and a half before the actual data were available for them to review, health care workers were told that they essentially had to administer MP to SCI patients. When the publication finally appeared in print, many physicians contested the importance of the findings, which they thought had been tremendously overstated. This bitterness about the way the NASCIS 2 results were initially disseminated probably accounts for much of the ongoing antagonism toward these trials.

Pearl

The manner in which the results of NASCIS 2 were released and, in many opinions, overstated no doubt contributed to the ongoing animosity that many clinicians feel toward the NASCIS investigations.

METHODOLOGY

NASCIS 2 attempted to randomize SCI patients within 12 hours of injury to one of three groups: placebo, MP, or naloxone. MP was given as a bolus of 30 mg/kg of body weight over 15 minutes, followed 45 minutes later by a continuous infusion of 5.4 mg/kg/h for 23 hours. The naloxone dose was 5.4 mg/kg for the 15-minute bolus, followed 45 minutes later by 4.0 mg/kg/h for 23 hours.

Neurologic assessment was divided into motor function, sensation to pinprick, and sensation to light touch. For motor testing, 14 muscle segments (innervated by 14 different levels of the spinal cord from cervical to sacral) were each given one of six scores: 0 (no contraction), 1 (reduced contraction), 2 (active movement without antigravity; i.e., side-to-side but not upward), 3 (active movement against gravity), 4 (reduced function but active movement against resistance), or 5 (normal function). Possible scores ranged from 0 to 70.

For sensory testing, 29 segments of the spinal cord from C2 to S5 were evaluated bilaterally for pinprick and light touch according to the following scale: 1, absent; 2, decreased; or 3, normal. Possible scores ranged from 29 (absent at all levels) to 87 (normal at all levels).

The primary end point was the change in neurologic score between baseline assessment and follow-up evaluation. Additional planned analyses would analyze the effects of the protocol, of the time the MP dose was received, and of the degree of neurologic loss (complete or incomplete).

Each patient was placed into one of five motor and sensory categories. The five motor categories were quadriplegic, paraplegic, quadriparetic, paraparetic, and normal. Sensory categories were as follows: analgesic and anesthetic at or above T1, analgesic and anesthetic below T1, hypalgesic and hypesthetic at or above T1, hypalgesic and hypesthetic below T1, and normal.

The analysis of neurologic scores used only the right side of body, and to simplify the presentation of results, only data from the right side were shown in the initial publication. However, when the analyses were performed on the left side, essentially identical results were obtained.

RESULTS

A total of 487 patients were randomized: 162 to MP, 154 to naloxone, and 171 to placebo.

Six-Week Results

At the 6-week outcome assessment of all patients in the MP and placebo groups, there was no significant difference in neurologic change scores between these groups. However, if only those patients treated within 8 hours were considered (remember that one of the a priori hypotheses was that effects of treatment would be influenced by how quickly the drug was given), the MP group showed significantly more improvement in motor and light touch scores, with a trend toward improvement in pinprick scores. The 8-hour cutoff interval was chosen because the average time from
injury to bolus dose for all patients in the entire study was 8.7 hours, with a median time of 8.5 hours. The exact number of patients treated within 8 hours drops to 66 MP patients and 69 placebo patients (62 and 67, respectively, at 6 months) out of the 487 entered into the trial.

Another a priori hypothesis was that any treatment effects would be influenced by the severity of injury. Accordingly, patients treated within 8 hours were classified as either plegic with total sensory loss, plegic with partial sensory loss, or parietic with variable sensory loss. At 6 weeks, the first group had more improvement in motor function than placebo, with strong trends for sensory improvement. The small number of patients in the second group demonstrated no differences between MP and placebo, whereas the third group exhibited a trend toward improvement only in motor score.

**Six-Month Results**

At 6 months after injury, the entire MP group showed significantly more improvement than the placebo group in pinprick and light touch scores, but apparently not in motor improvement. However, among patients treated within 8 hours of injury, the MP group demonstrated significantly more improvement than placebo in all three measures: motor, pinprick, and light touch.

At 6 months postinjury, MP patients who were plegic with total sensory loss had more improvement than placebo patients in all three neurologic measures. On the other hand, patients who were plegic with partial sensory loss showed the same amount of improvement in neurologic scores regardless of group. Among those who were parietic with variable sensory loss, only the amount of motor improvement was statistically significant.

Among patients who were treated after 8 hours, there were no differences in neurologic change scores related to treatment. Likewise, patients treated with naloxone demonstrated no statistically significant differences in improvement.

The above analyses were done on all randomized patients; that is, "intent-to-treat analyses." If only those who received drug within the protocol’s time limits are analyzed, the differences in favor of MP become larger.

If patients were classified as quadriplegic, paraplegic, quadriparetic, or paraparetic, and if they were treated within 8 hours, the odds of improving by a full category at 6 weeks tended to be higher in the MP group. The degree of such improvement was less evident at 6 months and was not seen if drug was given after 8 hours.

For MP, naloxone, and placebo patients, wound infections occurred in 7.1%, 3.3%, and 3.6% of patients, respectively, and gastrointestinal bleeding occurred in 4.5%, 2.0%, and 3.0%, respectively. These differences were not significantly different.

**NASCIS 2: One-Year Report**

The numbers of patients who received study drug within 8 hours of injury and who were assessed 1 year after injury were 62 for MP, 56 for naloxone, and 65 for placebo.

There were no significant differences in neurologic function by treatment group. However, if only those patients who received their drug bolus within 8 hours were considered, those in the MP group demonstrated significantly greater motor recovery (17.2 versus 12.0; p = .03). Improvements in pinprick and light touch were not statistically significant.

Among all patients who were randomized more than 8 hours from injury, there were strong but statistically nonsignificant trends for the MP and naloxone groups to recover less neurologic function than the placebo group.

**Pearl**

NASCIS 2 demonstrated that MP was associated with more motor recovery than placebo, but only if given within 8 hours of injury. Beyond 8 hours, administration of MP was associated with a nonsignificant trend toward recovery of less neurologic function than that seen in the placebo group.

Analysis of changes in neurologic function scores 1 year after injury in patients who received the study drug within 8 hours demonstrated significant improvement in motor scores for patients who were plegic with total sensory loss or parietic with variable sensory loss, but not in the small group of patients who were plegic with partial sensory loss. When patients were analyzed for improvement from an abnormal neurologic category (quadriplegic, paraplegic, quadriparetic, or paraparetic) to a higher level, there was a nonsignificant trend for more improvement with MP.

The groups did not differ significantly for any of 13 complications, most of which were reported at the 6-week follow-up. For all follow-up periods, only three complications approached statistical significance, and all occurred in comparisons of the naloxone group to the placebo group. The authors note that complications, even if related to the MP treatment, were manageable and were clearly outweighed by the potential benefits of improved recovery from SCI.

**NASCIS 3: Six-Month Report**

**Methodology**

NASCIS 3 randomized patients into three groups: 24 hours of MP (24MP), 48 hours of MP (48MP), or 48 hours of the free radical scavenger tirilazad mesylate (TM). Because of the presumed importance of ongoing
lipid peroxidation and hydrolysis as deleterious patho-
physiologic sequelae of SCI, it was thought that exten-
sion of the duration of MP therapy beyond 48 hours
might confer more benefit than a 24-hour period of
administration. In addition, it was thought that admin-
istration of the lipid peroxidation inhibitor TM might
facilitate neurologic recovery in a manner comparable
to that of MP, but with fewer complications.

Two preplanned subgroup analyses were early versus
late initiation of treatment within the 8-hour window
and the effect of treatment in patients with complete
versus incomplete neurologic function.

Patients in all three groups received a bolus of
30 mg/kg of MP, followed by randomization to one of
the three arms of the study. Neurologic assessment was
the same as that used in the previous NASCIS trials,
with the addition of deep pain and pressure responses
in the wrist, thumb, little finger, knee, ankle, and great
toe. These responses were scored as 1 (absent), 2
(decreased), or 3 (normal).

Patients were also assessed with the FIM, which
assesses self-care, sphincter control, mobility, locomo-
tion, communication, and social cognition. Overall
scores range from 18 (needing assistance in all areas)
to 126 (completely independent). The four outcome
categories of the FIM are complete dependence, modi-
fied dependence, modified independence, and com-
plete independence.

Analysis of results included summaries by degree of
neurologic loss (complete or incomplete) and by time
of drug bolus. Time of bolus was divided according to
administration less than or equal to 3 hours versus
greater than 3 hours from injury; 3 hours was used
because it was the mean and modal interval from
injury to bolus dose.

RESULTS
A total of 499 patients was evenly distributed ac-
ross the three groups. Although the TM group had signifi-
cantly worse motor function than the two MP groups,
the average motor function was not significantly dif-
f erent between the 24MP and 48MP groups.

Six-Week Results
By intent-to-treat analysis at 6 weeks, the 48MP group
demonstrated a nonsignificant trend toward more
motor improvement than the 24MP and TM groups.
The difference between 24MP and 48MP became sig-
nificant if 38 noncompliers were excluded from the
analysis.

Intent-to-treat analysis of the effect of timing of drug
bolus revealed significantly more motor improvement
at 6 weeks among the 48MP group compared with the
24MP group, a difference that became more signifi-
cant if only compliers were analyzed. When patients
were analyzed by presence of complete or incomplete
injuries, there were no significant differences at 6 weeks
between 24MP and 48MP patients, but statistical sig-
ificance was reached for both complete and incomplete
groups if only compliers were analyzed.

Six-Month Results
At 6 months, neither intent-to-treat analysis nor compli-
ers’ analysis revealed significant differences in motor or
sensory function between the entire 24MP and 48MP
groups. There were no differences among patients who
received the bolus within 3 hours, but among those
bolused between 3 and 8 hours, both intent-to-treat
and compliers’ analysis revealed significantly greater
change scores in the 48MP group compared with the
24MP group.

When 6-month improvement in completely and
incompletely injured patients was compared in 24MP
versus 48MP groups, the patients with complete injuries
demonstrated more improvement in the 48MP group,
a difference that just reached statistical significance.
This statistical significance was not seen among incom-
pletely injured patients. The same pattern was seen in a
compliers’ analysis: patients with complete injuries did
better in the 48MP group, but the difference in improve-
ment was not statistically significant among those with
incomplete injuries.

There were no significant differences in improve-
m ent of sensory function among the three groups.

When 6-month FIM scores were analyzed, intent-to-
treat analyses demonstrated significantly better scores
in the 48MP group (compared with the 24MP group) in
self-care and sphincter control, and these differences
remained significant when a compliers’ analysis was
performed. Total FIM scores were not significantly bet-
ter by intent-to-treat analysis, but they did reach a p
value of .05 when a compliers’ analysis was performed.

Finally, improvement in neurologic recovery from
one of the four abnormal categories (quadrilegic,
paraplegic, quadriparetic, or paraparetic) to a higher
category was examined. Six-month intent-to-treat
analysis revealed a strong trend toward such improve-
m ent in 48MP patients versus 24MP patients. When
only those who began treatment between 3 and 8 hours
were analyzed, the differences in improvement bec ame
statistically significant.

Improvements in the TM group did not reach statis-
tical significance. However, the comparability of motor
recovery rates in the TM and 24MP protocols was
stressed by the authors.

Complications
There were no differences in survival in the three
groups. Thirty-two categories of potential complica-
tions occurred with essentially the same frequencies in
all groups. An exception was severe pneumonia,
which occurred in 5.8% of 48MP patients, 0.6% of TM
patients, and 2.6% of 24MP patients (p = .02).
Pearl

In NASCIS 3, severe pneumonia was significantly more common in the 48MP group than in the 24 MP group. Overall mortality rates did not differ, but the 48MP group tended to have more deaths from pneumonia, respiratory distress syndrome, and respiratory failure than the 24MP group (six patients vs. one patient).

NASCIS 3: One-Year Report

At 1 year, there were no differences in mortality rates among the three groups. However, in terms of causes of death, the 48MP group exhibited a strong trend toward more deaths from pneumonia, respiratory distress syndrome, and respiratory failure than the 24MP group, although the numbers of patients are small (six patients versus one patient; p = .056).

By intent-to-treat analysis, there was a strong trend for the 48MP group to have a better motor change score than the 24MP group if treatment was initiated between 3 and 8 hours (19.0 versus 12.6; p = .053). This difference became significant if the analysis included only compliers (change score 19.4 versus 13.3; p = .032).

Intent-to-treat analysis revealed no significant differences in sensory change scores, and analysis based on the completeness of injury found no significant differences between groups with different degrees of baseline neurological function. By compliers’ analysis, patients with incomplete injuries and variable sensory loss demonstrated a strong trend toward greater motor improvement in the 48MP group compared with the 24MP group (31.4 versus 27.1; p = .054).

Improvement from one of the four abnormal categories (quadriplegic, paraplegic, quadriparietic, or paraparetic) to a higher category or to normal was not significantly different between groups. One-year FIM scores were not significantly different between groups.

New complications reported for the first time between 6 months and 1 year occurred with essentially equal frequency in the three groups.

The authors conclude that these findings are of benefit to the approximately half of SCI patients who are admitted in the 3- to 8-hour window after injury. However, the significantly greater incidence of pneumonia at 6 weeks and the possible increase in mortality rate from pulmonary problems mandate caution and suggest that patients be treated with the 24-hour protocol whenever possible.

Summary of the NASCIS Trials

The above review may seem complex and confusing, and many who attempt to review the NASCIS literature often end up feeling the same way. The essential findings are summarized below.

NASCIS 2

Outcomes at 1 year did not differ between MP and placebo groups when all patients in each group were compared. When only patients treated within 8 hours were analyzed, the MP group had significantly greater motor change scores (17.2 versus 12.0; p = .03), but no significant difference was seen in pinprick and light touch scores.

NASCIS 3

The 48MP group had a better motor change score than the 24 MP group at 1 year only if patients bolused between 3 and 8 hours were compared (19.0 versus 13.7); this difference was of borderline statistical significance (p = .053). If only patients who complied with the protocol were included, this difference became more significant (19.4 versus 13.3; p = .032). There were no significant differences in sensory change scores. FIM scores did not differ significantly between groups.

Severe pneumonia occurred significantly more often in 48MP patients (5.8%) than in 24MP patients (2.6%) and TM patients (0.6%) (p = .02). Overall mortality rates did not differ among the three groups, but in terms of causes of death, the 48MP group exhibited a strong trend toward having more deaths from pneumonia, respiratory distress syndrome, and respiratory failure than the 24 MP group (p = .056).

Some Criticisms and Defenses of the NASCIS Trials

Space does not permit a comprehensive review of all the published critiques of the NASCIS trials. Additional discussion is readily available in the literature. Some of the more frequently voiced criticisms are discussed below, along with possible responses that might be offered by the NASCIS investigators. Readers may judge for themselves both the validity of the criticisms and the soundness of the responses.

1. Criticism: It is not clear how the reported gains in motor recovery associated with MP relate to actual functional benefit because no measure of functional outcome was included in NASCIS 2. Is an improvement of five or six points in motor change score due to marked improvement in an important muscle group or only to meaningless twitches in two or three muscle groups below the level of injury? Response: NASCIS 2 was designed to measure only neurologic status, not total functioning of the patient. At the time that NASCIS 2 was being planned, no generally accepted, validated, and reliable scales existed to
measure activities of daily living, urinary function, catheterization, or ambulation. Neurologic function was used because those data have high interobserver reliability and relate to the function of specific spinal tracts. "The NASCIS group prefers to interpret the results of NASCIS 2 as showing that neurological recovery can be influenced by pharmacological interventions early in the injury process..." [italics present in original document].

The subsequent NASCIS 3 trial did use the FIM. Subsequently, the relationship between motor and 1-year FIM scores in NASCIS 3 was analyzed, and the results were used to estimate the degree of improvement in FIM scores that would be expected from the motor recovery observed in NASCIS 2. Overall, almost 10% of patients could be expected to show at least a 10% improvement in the modified FIM scores used in this analysis, an improvement that would be expected to represent a clinically important degree of recovery (class II data).^23

3. Criticism: Why are results presented only as "change scores"? Absolute scores should demonstrate the same trends if the data are reliable.

Response: Change scores are less variable than absolute neurologic scores and therefore permit more robust tests of the study hypotheses. Also, because neurologic change was the clinical outcome of interest in the trials, change scores were the most appropriate method of assessing treatment effect.\(^{13}\)

4. Criticism: For the subset of patients who improved, no baseline epidemiologic, neurologic, or physiologic data are provided, even though these are the patients of greatest interest and the patients upon whom the conclusions of the studies are based.\(^{5}\)

Response: Randomization ensured that severity of initial injury (as well as other factors) was distributed equally among groups, and any residual variability in neurologic injury was controlled for in the statistical analyses. This study was analyzed by well-accepted statistical techniques that withstood the process of peer review.\(^{13}\) "After-the-fact" analyses of only patients who demonstrated more improvement are scientifically invalid; stratification analyses based on response rather than on treatment violate fundamental principles of analysis of clinical trial data.\(^{15}\) Also, when analyses suggested that other patient characteristics might be influencing treatment effects (which occurred infrequently), they were controlled in multivariate analyses.\(^{12}\)

5. Criticism: Why was there no improvement in sensory function in the MP group?\(^{19}\)

Response: Sensory scores tended to improve more in MP patients, but the improvement was not statistically significant. Reasons may include the greater variability of measuring pinprick and light touch as compared with motor function, the lower resolution of a three-point sensory scale (as opposed to a six-point motor scale), and the possibility that MP does not have as much effect on sensory function as on motor function.\(^{15}\)

6. Criticism: Motor scores are reported only for the right side, but sensory scores are reported bilaterally.\(^{16,22}\)

Response: All three NASCIS investigations analyzed only data from the right side.\(^{1,3,9}\) Neurologic function between the two sides was highly correlated, and summing both sides would amount to measuring the same phenomenon twice. This practice would be anticonservative with respect to detection of therapeutic effects.\(^{7,23}\) In other words, using data from only one side actually made it harder to demonstrate that any improvement that occurred was associated with use of MP.\(^{26}\)

7. Criticism: A placebo arm was not used in NASCIS 3.\(^{18,22}\)

Moreover, it must be remembered that mean scores are obtained in an entire population of patients. Some patients may not benefit at all, but others may demonstrate a great deal more recovery than the mean.\(^{15}\) The likelihood of recovery is significantly improved in patients who receive MP after SCI, but this is not the same as saying that all patients will recover by the mean value.\(^{12}\)

2. Criticism: The 8-hour time limit was not specified as a hypothesis prior to commencement of NASCIS 2, and the division into less than 3 hours versus 3 to 8 hours in NASCIS 3 was not specified in advance.\(^{5,16,18,22}\)

Response: Stratification of data into patients who were treated "early" versus "late" was planned from the beginning. The cut-off time of 8 hours in NASCIS 2 was chosen because that was closest to the median time of 8.5 hours for bolus dose of the entire population, thus permitting adequate statistical power for testing the effects of "early" versus "late" administration of MP.\(^{7,13,24}\) In NASCIS 3, 3 hours was chosen for the same reason; that is, it was the median and mean time between injury and bolus.\(^{7,9,24}\) No time points other than the ones presented were analyzed.\(^{7,24}\)
Response: Randomizing patients to placebo when an effective therapy is available is unethical. Based on their findings in NASCIS 2, the NASCIS neurosurgeons felt it would be unethical to include a placebo group in NASCIS 3.

8. Criticism: In NASCIS 3, the two steroid arms were not equivalent: 24.7% of patients in the 24MP group had normal motor function versus 13.9% in the 48MP group (p = .012). Because the treatment arms are unbalanced, the small gains attributed to 48 hours of MP treatment are questionable.18

Response: Normal patients are uninformative because they have no room for improvement and thus do not contribute to a change score (unless they worsen, which was very rare). Also, a correct baseline analysis would account for all levels of severity, and such an analysis reveals much greater equivalence between the MP groups in NASCIS 3.12

9. Criticism: The detailed neurologic examination must take an hour (by conservative estimate). Pain, anxiety, distractions, fractures, and other accompanying injuries, etc. must have prevented many patients from completing the entire assessment.18

Response: In NASCIS 3, only 2.5% of exams were incomplete, and in these cases data from the other side of the body were used. For NASCIS 2, retrospective analysis suggests an incompletion rate of 4.8 per 10,000 required parameters.12

10. Criticism: NASCIS 2 has not been replicated. The authors cite a Japanese article in a journal not indexed on Medline as a description of a study that replicates NASCIS 2,7,12,24,25 but the study described in that article did not utilize blinding, did not use a placebo, compared groups with different baseline motor and functional scores, and gave MP to approximately a fifth of the control group.18 The NASCIS investigators also refer to a relatively small French study in a manner that implies that the results of that study are supportive of the NASCIS results,7,12,24,25 but the abstract of that study (which is available in English on Medline) concludes that there is no evidence of benefit of medical treatment in SCI (class II data).27

Response: A meta-analysis that includes these trials (supplemented as needed with additional data obtained from the authors of the studies) confirms the efficacy of MP in SCI.11 Other studies on whiplash and microdiskectomy also support the use of MP.12

11. Criticism: MP treatment is associated with a very high complication rate.16,18,22

Response: In NASCIS 2, frequency rates for 18 complications were comparable in MP and placebo groups.12 In NASCIS 3, 31 and 50 patients, respectively, would have to be treated with 48 hours of MP in order for a single case of severe pneumonia or severe sepsis attributable to therapy to occur.7 The availability of effective treatments for these conditions compared with the lack of effective treatment (other than MP) for SCI suggests that the risk–benefit ratio favors MP. Review of all deaths in NASCIS 2 and NASCIS 3 revealed no direct association of any individual patient’s death with his or her drug therapy, and overall mortality was essentially the same in all treatment groups.7 A recent meta-analysis of 51 trials of administration of a high dose of MP in elective and trauma surgery found no significantly increased risk of gastrointestinal bleeding, wound complications, pulmonary complications, or death (class III data).26 In fact, MP was associated with a significantly lower rate of pulmonary complications, primarily in trauma patients.

12. Criticism: A “compliers’ analysis” is not valid in NASCIS 3, and the fact that dropping the small number of noncompliers has such a dramatic effect on p values in NASCIS 3 indicates that the data are “unstable.”16

Response: Although the primary analysis in the NASCIS trials was intent-to-treat, a compliers’ analysis is also of interest because it more likely reflects the effects of a therapy in ideal circumstances. Furthermore, one cannot infer anything about the direction of effect of individual patients in a trial.7

13. Criticism: In NASCIS 3, analysis of 1-year motor change scores in the 48MP versus 24MP groups (19.0 versus 13.7, respectively) yields a p value of .053, which is not significant.16,22

Response: The p value is quite close to the significance threshold of .05, and compliers’ analysis reveals that the difference between groups becomes significant (19.4 versus 13.3, p = .032).10 The potential for improvement in these patients with otherwise untreatable injuries outweighs the low risk of MP-associated complications, which are usually minor.

14. Criticism: Statistical analysis in the NASCIS trials is inappropriate because of multiple comparisons. So many comparisons were performed among different subgroups that there is a good likelihood that several comparisons would be statistically significant by chance alone.16,18 Correction of analyses to minimize random associations from appearing statistically significant just by chance should have been used.16

Response: Many statisticians consider such types of statistical correction to be overly simplistic and even inappropriate for analysis of related aspects of a common clinical phenomenon; for example, motor function, pinprick sensation, and touch sensation.7 Furthermore, of the 54 comparisons in NASCIS 2, statistical significance was reached for none of them, whereas statistical
significance would be expected for only two or three by using the statistical correction technique suggested by some critics. The probability that nine comparisons would be significant by chance alone is .0013. Four additional comparisons had p levels between .06 and .08, further strengthening the evidence supporting the use of MP.7

15. Criticism: The second a priori hypothesis of NASCIS 2—that treatment effect would be influenced by the severity of injury—was not addressed for the entire population of study patients, but only for patients who had been already stratified according to the first hypothesis, that is, time of MP administration.22

Response: This technique is a perfectly valid analytical strategy that documents the effects of recovery while, at the same time, analyzing the comparisons of principal interest: drug effects within severity groups.28 Randomization within subgroups is unnecessary because of the risk of unblinding when randomization is conducted within smaller groups and because a patient may belong to several groups, making randomization within all of them impossible. Thus, postrandomization stratification, which was performed in the NASCIS investigations, is the most common analytic strategy. Importantly, such stratifications must be part of the original proposal, which was the case in the NASCIS trials.12

16. Criticism: No standardized medical or surgical management protocol was used, and possible effects of aggressive medical or surgical management on outcome were not addressed.22

Response: Standardization of treatment regimens is not common in clinical trials, not possible in this type of study, and not necessary. Randomization allows for control of variability of concurrent therapies by distributing any variations in management evenly between groups. Moreover, when other treatments are provided in accordance with local practices, generalizability of trial results is enhanced. Finally, published analyses have reported on the timing of surgery in NASCIS 2 and have documented that surgery did not affect the results regarding use of MP (class II data).35,29

Pearl
Use of MP in patients with gunshot wounds to the spine was not studied in the NASCIS trials. Other investigators have suggested that it confers no benefits and may contribute to the development of complications (class III data). Other common uses of MP, such as for pediatric SCI, prophylaxis during high-risk spine surgeries, etc., were not addressed in the NASCIS trials and should be considered to be level III recommendations at best.

Conclusions
NASCIS 2 demonstrated that 24MP treatment is associated with more neurologic improvement than placebo. Unfortunately, despite a great deal of hype and hoopla that preceded the actual publication of the study, the average amount of improvement is modest. The manner in which the results were released and, in many opinions, overstated no doubt contributed to the ongoing animosity that many clinicians feel toward the NASCIS investigations. Furthermore, in NASCIS 3, the motor benefits of 48 versus 24 hours of MP were, at best, of borderline statistical significance (p = .053).

Many criticisms focus on the clinical relevance (or lack thereof) of the relatively minor amount of average neurologic improvement. Nevertheless, this improvement (however slight it may appear) was significantly greater in the MP group. The overall lack of major side effects indicates that the risk:benefit ratio is acceptable.

Other criticisms focus on the analytic and statistical methods of the studies. The NASCIS authors generally do a good job of refuting these criticisms. However, the NASCIS authors also use some weak arguments and irrelevant references to defend their studies; for example, citing trials of MP efficacy in whiplash and lumbar discectomy.12 Such arguments make the NASCIS authors appear almost desperate to find any kind of support for the use of high-dose steroids.
In their responses to various critics, the NASCIS authors also tend to put too much favorable spin on results that are not significant, and they gloss over major negative results of the NASCIS trials. These tendencies may perpetuate confusion and, in some cases, contribute to a perceived lack of credibility of the results. It would be better if the authors did not try so hard to "sell" their studies and, instead, let the results (both strong and weak) speak for themselves.

Recommendations

High-dose methylprednisolone after SCI does appear to improve neurologic function more than placebo. However, the improvement in neurologic change scores has not been shown to translate into improved function in SCI patients. For these reasons, use of MP as recommended in the NASCIS trials should be viewed as a level II recommendation.

Acknowledgment

This chapter is based in part on material presented at the meeting "Trauma and Critical Care—2001," Las Vegas, Nevada, March 2001.

References