A Review and Methodologic Critique of the Literature Refuting Whiplash Syndrome

Michael D. Freeman, D.C., Ph.D., M.P.H., Department of Public Health and Preventive Medicine, Oregon Health Sciences University School of Medicine, Portland, Oregon.

Arthur C. Croft, D.C., M.S., Spine Research Institute of San Diego, San Diego, California

Annette M. Rossignol, Sc.D., Department of Public Health, Oregon State University, Corvallis, Oregon

David S. Weaver, D.C., Private Practice, Keizer, Oregon

Mark Reiser, Ph.D., Department of Economics, Arizona State University, Tempe, Arizona

Address reprint requests to:

Dr. Michael D. Freeman, 4747 River Road North, Salem, Oregon 97303 USA, phone 503-393-3133 fax 503-463-5042

Corresponding author:

Dr. Michael D. Freeman, 4747 River Road North, Salem, Oregon 97303 USA, phone 503-393-3133 fax 503-463-5042

Precis

The purpose of the present study was to evaluate the methodology of the literature refuting whiplash syndrome. Over 2000 papers in the whiplash literature were reviewed for publications that clearly refuted the validity of whiplash syndrome. This literature search revealed 20 such papers. These papers were subsequently reviewed for methodologic flaws that may have invalidated their conclusions. All 20

papers were found to have significantly flawed methodology, and it was determined that their conclusions regarding whiplash syndrome were not supported by their research methods.

Unstructured Abstract

The validity of whiplash syndrome has been a source of debate in the medical literature for many years. Some authors have published papers that suggest that whiplash injuries are impossible at certain collision speeds, others have stated that the problem is psychological, or a result of secondary financial gain. These papers contradict the majority of the literature, which shows that whiplash injuries and their sequelae are a highly prevalent problem that affects a significant proportion of the population. The authors of the current literature critique reviewed the biomedical and engineering literature relating to whiplash syndrome, searching for papers that refuted the validity of whiplash injuries. Twenty papers containing nine distinct statements refuting the validity of whiplash syndrome were found that fit the inclusion criteria. The methodology described in these papers was evaluated critically to determine if their observations regarding the validity of whiplash syndrome were scientifically sound.

The authors found that all of the included papers contained significant methodologic flaws with regard to their statements refuting the validity of whiplash syndrome. The most frequently found flaws were inadequate study size, non-representative study sample, non-representative crash conditions (for crash tests), and inappropriate study design. As a result of the current literature review, it was determined that there is no epidemiologic or scientific basis in the literature for the following statements: whiplash injuries do not lead to chronic pain, rear impact collisions that do not result in vehicle damage are unlikely to cause injury, whiplash trauma is biomechanically comparable to common movements of daily living, among others.

Introduction

One of the more frequently disputed conditions in the medical literature in recent decades is the constellation of symptoms comprising acute whiplash, and its chronic iteration, late whiplash (collectively known as whiplash syndrome). The primary reason for the dispute stems from the fact that the validity of whiplash syndrome often is a key issue in litigation arising from the alleged etiology of the whiplash; i.e. a motor vehicle crash in which the injured party is not at fault. The judge and/or jury in such cases are asked to weigh opposing medical and scientific evidence supporting both the plaintiff’s position that whiplash injuries and their sequelae are real and the defense position that the injuries are manufactured or greatly exaggerated. Over $29 billion per year is spent on whiplash injuries and litigation in the United States alone ().

It is not surprising, considering the financial stakes, that many medical experts have dedicated their professional careers to one side or another of the whiplash controversy. These experts increasingly are relying on medical and engineering literature to support both sides of the debate over the validity of whiplash syndrome.

A recent review of the literature reported over 10,000 articles relating to whiplash injuries (). The majority of this literature is devoted to probing fundamental questions about whiplash injuries, such as mechanism of injury, pathogenesis, and epidemiology. Over 30 epidemiologic studies have been published that document the cumulative incidence (risk) of chronic (lasting longer than six months) whiplash symptoms, or "late whiplash." In a recent publication, thirteen of these studies were considered sufficiently well constructed (low selection bias, sufficient study size, adequate research methodology) to be relied upon for an accurate clinical projection for late whiplash (). A study population-weighted meta-analysis of these studies reported a 0.33 risk of late whiplash at 33 months post-injury for those seeking treatment for acute whiplash injuries (1). Thus, the epidemiologic literature appears to support a substantial risk of chronicity following acute whiplash injury.

Federal government statistics and epidemiologic studies indicate that whiplash syndrome affects a large number of people. The National Highway Traffic Safety Administration reports that, in 1995, there were 5,500,000 Americans injured in motor vehicle crashes (MVC) (). A large, population-based study found that 53% of MVC injuries include whiplash injuries, amounting to 2,900,000 acute whiplash cases in 1995 (), or an incidence rate of 1,107 per 100,000 person-years (1). If, as is suggested by the results of

the meta-analysis described earlier, 33% of acutely injured persons continue to experience symptoms at 33 months, then as many as 900,000 new cases of late whiplash may have occurred in the U.S. in 1995.

A recent case-control study of 665 subjects with chronic spine pain found that 45% of patients who reported having at least one intrusive episode of neck pain weekly for more than six months attributed the onset of their symptoms to a whiplash injury. While it is important to keep in mind that the results of any case-control study must be interpreted carefully due to the potential effect of recall bias, if the results of this chronic neck pain study are applied to what, to the authors' knowledge, is the most conservative published estimate of the prevalence of chronic neck pain in the population (13.8%) (), then it can be reasonably, if cautiously, estimated that 6.2%, or 15.5 million Americans currently have late whiplash. Other authors have estimated chronic neck pain prevalence to be as high as 32.9% for women and 27.5% for men (); therefore, the prevalence of late whiplash could be substantially higher.

Despite the strong epidemiologic evidence supporting whiplash syndrome as a valid clinical entity that leaves many persons with permanent symptoms, numerous papers have been published, the majority since 1990, that refute the validity of some or all aspects of whiplash syndrome. And, while the entire whiplash literature base has been criticized for methodologic weakness in general (2,,), the quality of the literature refuting whiplash syndrome has stood largely unchallenged.

The present study reviews, from a methodologic perspective, the literature refuting whiplash syndrome. The objectives of this review is twofold. The first objective is to determine whether there are significant methodologic flaws in the individual papers that may undermine the accuracy of their conclusions regarding the biomechanics, pathogenesis, or epidemiology of whiplash syndrome. The second objective is to determine, if there are methodologic flaws in the literature, whether there are categorical flaws that are common to more than one study.

Methods

The literature was searched for papers that contained statements in the abstract or conclusions that refuted the validity of part or all of whiplash syndrome. For the purpose of the present study, whiplash syndrome was defined as injuries and their sequelae resulting from indirect trauma to the spine, following low to moderate severity motor vehicle crashes. Late whiplash was defined as whiplash syndrome persisting for greater than six months.

The literature was searched for titles or abstracts containing the term "whiplash." Literature databases searched were Medline, SAE, IRCOBI, and NTIS for the years 1966 through 1997, in addition to published studies the authors were aware of that contained statements refuting whiplash syndrome.

Over 2000 papers were reviewed at least cursorily to determine relevance to the current review. Of these, more than 700 of the most relevant papers were read in extenso. The articles were reviewed for specific statements that were considered to be contrary to the authors' understanding of how the majority of the current literature characterizes the biomechanics, pathogenesis, and epidemiology of whiplash syndrome. These statements were categorized and described. In addition, logical implications of the statements that may arise in a medico-legal setting were extrapolated and described. The statements and their respective implications are listed in Table 1.

The studies then were reviewed by the authors for the presence of significant methodologic flaws. A significant methodologic flaw was defined as a potential threat to the validity of the study in light of the study's conclusions regarding whiplash syndrome. In other words, while some of the study's methods and inferences regarding whiplash syndrome may be valid, the study methods were evaluated solely in reference to its conclusion or conclusions that caused it to be included in the present critical review.

The authors were asked to critique the articles individually, and if methodologic flaws were found, to describe them. The methodologic errors were then described, categorized, and put into table form (see Table 2).

Results

The literature search revealed 20 papers containing statements in the abstract, conclusions, or elsewhere in the text, that were interpreted as refuting whiplash syndrome. Those statements are summarized in Table 1 at the end of the results section.

The papers ranged, with respect to study type, from literature reviews to cohort studies. The papers were either designed a priori as a refutation of whiplash syndrome, or they were designed for another purpose but made extrapolative statements that refuted the validity of whiplash. The papers were divided primarily between biomedical studies and editorials, and engineering studies.

All 20 papers were found to have significant methodologic flaws relative to their proclamations regarding the validity of whiplash syndrome. These flaws were of sufficient magnitude to cast doubt upon the theoretical basis for the stated link between the study results and the conclusions of the study regarding the validity of whiplash syndrome. The papers are categorized below, according to study type. A brief description of the major points of the paper is given, followed by a discussion of the methodologic flaws that were found in this review. If there were flaws that were common to more than one study in a category, then all of the papers with the common flaws are listed, followed by a description of the flaws.

COHORT STUDIES

1. Schrader et al. studied 202 individuals in Lithuania who had been involved in a motor vehicle crash. This cohort was age and gender matched with a control group of 202 individuals who had no history of a MVC. The two groups were surveyed for neck pain an average of 21.7 months post-crash (relative to the time of the motor vehicle crash for the MVC-exposed cohort) and were found to have the same prevalence of neck pain. The authors concluded that whiplash injuries do not cause chronic symptoms, and the reason that late whiplash exists in industrialized countries is because insurance settlements are available to those claiming chronic pain ().

Methodologic Errors

Inadequate Sample Size This study was criticized because only a very small proportion of the exposed cohort (15% [31 subjects]) had been injured initially, and thus exposed to the putative etiologic agent in late whiplash (an acute whiplash injury) (). For the purposes of the current literature critique, a post-hoc sample-size calculation was performed on the data in this study, using an alpha of 0.05 and a beta of 0.20. The smallest detectable difference between the groups was 14.6%. Thus, 94% of the acutely injured subjects (29 of 31) in this study would have had to develop chronic symptoms to enable the authors to detect a statistically significant difference between the two groups, an extremely remote possibility. A recalculation of sample size using a meta-analysis-based estimate of effect (expected proportion chronic) of 5% (1) (that is, 33% of the 15% acutely injured subjects) demonstrates that the total study cohort needed to be at least 3000 in order to have sufficient statistical power to discern a significant difference between the two groups.

2. Balla reported on a cohort of 20 whiplash patients presenting to an orthopedist in Singapore with follow-up of more than two years (). He reported that none of the 20 patients had symptoms of late whiplash, and concluded that late whiplash was rare in Singapore, in comparison with a group of 300 Australian patients with late whiplash. Balla attributed the late whiplash rate difference between the two countries to cultural differences and economic factors, among others.

Methodologic Errors

Inappropriate Study Design Balla compared a group of 300 late whiplash cases to 20 subjects who had been evaluated following a whiplash trauma. Not only were the numbers in the two groups grossly disparate, the subjects were enrolled in two different studies using different enrollment criteria and study protocol. The 300 Australian subjects were selected for study because they had late whiplash. The 20 Singaporean subjects were recruited from a specialist’s practice on the basis that they had sustained an acute whiplash injury. As a result of different selection criteria for the two groups, and other

dissimilarities, the study could not validate or invalidate the author's hypothesis that the natural history of whiplash injuries in Australia differs from that of Singapore.

Inadequate Sample Size Twenty subjects is not a sufficient size for a prospective study of late whiplash. Using Balla's Singapore data, a post-hoc power calculation was performed, assuming that the risk of late whiplash in Australia at 33% (a literature based assumption) was an unlikely eight times greater than in Singapore. At least 44 randomly selected subjects would be needed in Singapore for such a study. Recalculation of power using a more reasonable risk ratio of three to one results in the need for 64 randomly selected Singaporean subjects. Our power calculation assumed several study factors not actually present in Balla's study; identical selection criteria in both countries, random subject selection with control for potentially confounding differences between the countries not attributable to cultural differences, and identical subject appraisal criteria.

Selection Bias Selection bias was introduced in this study when subjects in Australia were selected for study retrospectively based on their disease status (they already had late whiplash when the study was begun) and the subjects from Singapore were selected prospectively based on their exposure status (an acute whiplash injury).

3. Heise et al. reported on 155 patients presenting to an emergency room following a whiplash trauma. The patients were divided into two groups; 63 patients with (unspecified) radiographic evidence of cervical musculoskeletal injury, and 92 patients without radiographic evidence of injury. The two groups were examined and interviewed for TMJ symptoms at the time of initial presentation, then followed-up by phone interview one month and one year subsequently. The follow-up rate at one year was 70% of the positive radiographic findings group, and 65% of the negative radiographic findings group. None of the patients who were contacted at one year had continued symptoms of TMJ dysfunction. The authors concluded that the incidence of TMJ injury following whiplash trauma was "extremely low." ()

Methodologic Errors

Inappropriate Study Design The authors do not state their rationale for stratifying their cohort into two groups on the basis of "positive radiographic findings" of whiplash, which are unspecified. The authors of this review were unable to find any reference in the literature to a correlation between TMJ injury and radiographic findings of whiplash injury that would justify the study design employed by Heise et al..

Inadequate Sample Size Using a literature based estimate of effect of 0.04 (5) (that is, 4% of the whiplash-injured population will sustain a TMJ injury) an alpha of 0.05 and a beta of 0.20, we performed a post-hoc power calculation on Heise et al.'s data. Assuming only double the frequency of TMJ injury in the exposed group, the authors would have needed over 2,500 subjects for their study. Assuming a highly unlikely eight times greater frequency of TMJ injury between the two groups studied, the authors still would have needed over 650 subjects, four times greater than the number in the study.

CASE SERIES STUDIES

Spitzer et al., in their Quebec Task Force (QTF) on Whiplash-Associated Disorders monograph, conducted a retrospective case series study and a literature search, and issued a set of guidelines and recommendations based on the results. Among other things, the QTF concluded that whiplash injuries were "short-lived," involving "temporary discomfort," that the pain resulting from whiplash was "not harmful," and that whiplash injuries have a "favorable prognosis." They also concluded that 87% and 97% of their cohort "recovered" from their whiplash injuries at six months and 12 months post-crash, respectively (2).

Methodologic Errors (3)

Improper use of terminology The Results and Discussion section of the case series study contained numerous references to the percentage of the study population "recovered" at the time of cessation of compensation. However, the QTF did not gather any data regarding the symptoms, amount or type of treatment, or functional impairment of their cohort -- all factors necessary to determine the level of

recovery following an injury. The QTF chose to define "recovery" unconventionally as cessation of time-loss compensation. Not surprisingly, the QTF found that 87% and 97% of their cohort was "recovered" at 6 and 12 months post-crash, respectively. To refer to these individuals as recovered is misrepresentative of the data collected.

Unsupported conclusions In a table labeled "Prevalence of symptoms at follow-up," the QTF enumerated the four studies on prognosis that were accepted for review, along with their findings, which were as follows: Norris and Watt (1983) reported that 66% of their cohort had neck pain at an average of 2 years post injury (); Radanov et al. (1991) reported that 27% of their cohort were symptomatic 6 months post-crash (), and in a study 2 years later (1993), reported that 27% of their cohort had headaches 6 months post-crash (); and Hildingsson and Toolanen (1990) reported 44% of their cohort symptomatic at an average of 2 years post-crash ().

Yet based on their literature review and their cohort study, the QTF concluded that "Whiplash-associated disorders are usually self-limited," and "Patients should be reassured that most WAD are benign and self-limiting," inaccurately summarizing the results of their literature review and case-series study.

CROSS-SECTIONAL STUDY

Bovim et al., in their study of chronic neck pain in the general population in Norway, stated that 13.8% of respondents had "troublesome neck pain" for longer than six months. The authors compared this proportion to similar figures reported by previous authors regarding the risk of late whiplash following an acute whiplash injury and concluded that "chronic neck pain after whiplash injuries may be a continuation of pre-existing complaints (6)."

Methodologic Error

Misquoting Literature/Selecting Biased Literature The basis for the primary conclusion of the Bovim et al. paper is the comparison of their survey results to a literature-based estimate of the prevalence of late whiplash among the population of individuals who have sustained whiplash trauma. The authors referenced four papers that contained estimates of chronicity following whiplash. One of the papers, written in Norwegian, could not be evaluated for this critique. The remaining three papers were stated to have reported a prevalence of chronicity of 12-18%. However, the authors did not reference 27 of the 30 papers on whiplash prognosis available in indexed journals at the time of their study. A meta-analysis of the 13 highest quality papers on whiplash chronicity reported that 33% of whiplash-injured individuals will have chronic neck pain at 33 months post-crash (1). This more accurate appraisal of the literature-based estimate of chronicity invalidates Bovim et al.’s hypothesis that late whiplash is merely a continuation of preexisting neck pain. Additionally, Bovim et al. misquoted one of the papers, by Gotten, claiming 12-18% chronic, when in actuality, Gotten reported a prevalence of late whiplash of 46% at 12 months post-crash ().

CORRELATIONAL STUDY

Mills and Horne compared the rate of whiplash injuries in Victoria, Australia, to the rate in New Zealand. They reported that the rate was substantially higher in Victoria and concluded that the difference was attributable to the fact that an injured occupant in Victoria must seek compensation through the common law system, as opposed to New Zealand, where apparently it is less difficult to gain compensation for motor vehicle crash-related injuries. The authors concluded that Victorians are "more conversant with and more attuned to receiving compensation for injury, which may in itself be stimulus for claiming an injury that they would not normally have claimed for ()."

Methodologic Error

Unsupported Conclusions The authors do not present any evidence that supports their statement that the greater barriers to claiming compensation in Victoria actually increase claims of whiplash injury. Indeed, the logical conclusion is quite the opposite. The difference in the whiplash rate between Victoria and New Zealand may be accounted for by any of a variety of potentially confounding factors that may

exist between the two countries, including different criteria for reporting and recording whiplash injuries, different driving conditions, or different diagnostic classification systems.

LITERATURE REVIEWS/EDITORIALS

1. Ferrari and Russell, in their editorial/literature review, stated that over "2000 runs of volunteer collisions have been conducted using specialized sled devices and actual vehicles (old and new, big and small), and never, ever, has the multitude of chronic symptoms of whiplash patients been reproduced ()."

The authors stated that it is "unacceptable, however, to claim that a muscle sprain or some as yet unidentified injury is responsible for the chronic pain and the large number of symptoms of whiplash patients. Instead, the symptom complex can be explained as a whole not by an injury, but rather by a psychological disorder."

Methodologic Errors

Unsubstantiated/Unreferenced Claims Ferrari and Russell provide no citation for their statement regarding the number or scope of crash testing. The literature review performed for the present critique revealed published accounts of fewer than 100 volunteers in crash tests, with the largest single majority (42 subjects) from one study that was published after Ferrari and Russell published their paper (). Although the authors state that no crash test study has ever produced chronic symptoms, there is no evidence in the literature to substantiate this statement. The authors of the present critique were only able to find two studies with a total of nine volunteers that informally followed the subjects for more than a few days to determine if there were chronic symptoms following crash testing (,).

The authors do not cite any references to substantiate their statement that it is "unacceptable" to claim an as-yet unidentified cause of chronic pain following whiplash. While the authors state that no cause has been identified for chronic pain following whiplash, they ignore the research of Barnsley et al., who have quite convincingly demonstrated the cervical zygapophyseal joints as the origin of a substantial proportion of chronic neck and head pain following whiplash trauma (). Ferrari and Russell do not cite any references that substantiate their claim that late whiplash is a psychogenic illness.

2. In his literature review/editorial, Awerbuch stated that as soon as a doctor makes a diagnosis of whiplash, he or she is contributing to the patient's potential for chronicity. The author continued, "later the patient may be referred for a range of imaging (plain x-ray, computed tomography, isotope bone scan, MRI, or thermography) which can only be interpreted by the patient as being necessary to define the gravity of the 'whiplash' injury," thus, further contributing to the potential for chronicity ().

Methodologic Error

Unsubstantiated/Unreferenced Claim The author does not cite any published sources to substantiate the statement that treatment and diagnosis contribute to potential chronicity of whiplash symptoms. Awerbuch overlooks the alternative explanation that symptomatic patients may be more likely to need additional treatment and diagnostic testing.

CRASH TEST STUDIES

1. McConnell et al. (1993) reported the results of human volunteer rear-impact crash testing of four subjects. They determined that, in reference to whiplash injuries resulting from rear impact collisions, the threshold of a "very mild, single event musculoskeletal cervical strain injury" is a delta V (the absolute velocity change of the struck vehicle as opposed to the speed of the striking vehicle at impact) of four to five miles per hour ().

2. McConnell et al. (1995) studied the movements and acceleration forces sustained by seven human occupant volunteers subjected to repeated rear-end collisions of up to 6.8 mph delta V. They concluded that at a delta V of five mph "the likelihood of transient acute neck and shoulder muscle strain injury and possible mild compressive irritation of the posterior neck may increase" for the average vehicle occupant. They also concluded that any injury to the low back is "quite unlikely as a result of a low velocity rear end collision (23)."

3. West et al. studied the acceleration forces sustained by six human volunteers in crash testing of five different vehicles. They concluded that vehicle occupants are unlikely to be injured in collisions with an equivalent barrier speed (EBS) of less than eight miles per hour (EBS is an estimate of impact speed based on vehicle damage, compared to a known amount of damage from a 30 mph collision with a fixed barrier). The authors also stated that they did not observe jaw opening during crash testing and that this finding rebutted claims that TMJ injury can result from whiplash trauma ().

4. Szabo et al. (1994) reported on human volunteer crash testing of five subjects who were in vehicles that were struck in the rear at approximately 10 miles per hour by another vehicle, resulting in an average delta V of five miles per hour (23). The subjects were evaluated by an orthopedic surgeon and given an MRI scan before and after the crash testing. Although four out of five volunteers complained of headache directly following the crashes, none had symptoms that lingered for more than two days, and no subjects reported further symptoms during the following year. The authors concluded that rear-end collisions with a delta V of five mph or less were within human tolerance levels, and that injury was unlikely following such a collision. Szabo et al. concluded that the jaw does not open during whiplash trauma, and stated that their study results support an earlier author's contention that there is no potential for TMJ injury as a result of a whiplash trauma.

5. Szabo and Welcher (1996) reported on volunteer crash testing of four men and one women (). The subjects were each exposed to two rear-end collisions with an average closing speed of 8.9 mph and an average delta V of 5.8 mph. The authors concluded that "a rear impact with a change on velocity of [5 mph] or less is within tolerance for a reasonably healthy occupant…."

6. Mertz and Patrick (1967) studied the responses of a human volunteer, a cadaver, and anthropomorphic dummies to simulated rear-end collisions. They compared the responses of the volunteer to an index of neck injury that was developed for the study by statically loading the neck of one of the authors with tension to the point that the author felt that injury might occur. The authors concluded that a 10 mph rear-end impact for an unsuspecting occupant was within human tolerance for injury ().

7. Mertz and Patrick (1971) used an anthropometric dummy, four cadavers and one of the authors for sled testing simulating a rear-impact collision. The volunteer sustained accelerations at the head of 1.9-6.8 g with no injury. However, a 9.8 g acceleration resulted in both back and neck injury. The authors developed a guide for tolerance to injury in a whiplash trauma ().

8. Rosenbluth and Hicks studied the acceleration forces sustained by two human crash-test volunteers who were seated in a vehicle that was struck from behind at an equivalent barrier speed (EBS) of up to 4.8 mph. They concluded that an EBS of 4.8 mph was below the threshold of human injury tolerance. The authors also measured the acceleration forces at the head (as measured by tri-axial accelerometers affixed to a helmet) of a seven year-old child and a 29 year-old adult skipping rope. They reported that acceleration at the head was similar to that found in the crash testing ().

9. Howard et al. (1995) studied the acceleration forces at the TMJ that occurred during rear-impact crash testing of four human volunteers. Howard et al. used accelerometers fitted to a bite plate to measure the acceleration forces at the approximate level of the TMJ during 5 mph delta V impacts. They concluded that the forces measured at the jaw during crash testing constitute a "minor fraction" of the normal forces experienced during mastication, and that low velocity whiplash trauma cannot cause injury to the TMJ ().

10. Castro et al. studied the effect of 17 rear impacts with an average delta V of 7.1 mph on 14 men and 5 women (the authors did not specify which two subjects were excluded from crash testing) (). Of the 17 impact-exposed subjects, five (29%) complained of whiplash symptoms following testing, including one male subject who had objective findings of injury 10 weeks post-crash. The authors concluded that "the 'limit of harmlessness' for stresses arising from rear-end impacts with regard to the velocity changes lies between [6.2 mph] and [9.4 mph]."

Methodologic Errors

Inadequate Study Size (papers 1-10) When attempting to study a population sample, in order to make an inference that is applicable to a population beyond that of the study, it is essential to use inferential statistics to determine if the study results were causally related to the variables under study, or if they were due to random variation. With crash testing, the dependent variable (the variable under study) is injury status; either an occupant is injured or not injured. Because the two outcomes are mutually exclusive, a 95% confidence interval can be established for the study results using a binomial probability distribution that is based on the study size. That is, if the study were to be repeated, the 95% confidence interval tells us how many and how few injuries are possible, based on the results of the current study. The width of a confidence interval is indirectly related to the number of subjects in a study, because random error makes the interpretation of the study results less precise, e.g. if a coin is tossed three times and heads is observed all three times, it is much less precise to state that the coin has heads on both sides, in comparison with 100 coin tosses resulting in heads.

Even with crash testing with as many as 20 subjects who sustain no injury in the crash test, the probability of injury in a larger population is still 0.15 (based on the confidence interval), which means that three subjects could be injured the next time the same study is conducted with the same subjects, and those results would still be consistent with the results of the current study. Thus, the confidence interval for crash test studies of five or six subjects is too wide to conclude that no injury is possible under similar conditions. In order to adequately describe the range of injury responses for the general population, given the wide variety of human susceptibility to injury, vehicle types, crash conditions, etc., many hundreds, or even thousands of subjects would need to be studied in crash tests.

Non-representative Study Sample (papers 1-10) The subjects in the crash test studies consisted of the authors of the studies, employees of the corporations financing the study, and other associates of the authors who may have a vested interest in the outcome of the study. In addition, almost all of the test subjects were male. In order to generalize the results of any study to a larger population (in this case, the general population at risk for whiplash injuries) the study population must adequately represent the larger population.

Non-representative Crash Conditions (papers 1-10) Even if the numbers of subjects were sufficient to generalize the results of the above listed crash tests to the general population, the results would only be applicable perfectly healthy males who were prepared for a rear impact and perfectly situated in the vehicle seat at the time of impact. Only a very small proportion of the crash-injured population fits this description.

For their crash test, Mertz and Patrick (1971) used a sled seat with a specially designed head restraint that did not allow for any posterior movement of the head (see figure 1). The results of such crash testing are not generalizable to the population at risk for whiplash trauma, because car seats allow for posterior excursion of the head, which is the most significant injury-producing phase of whiplash trauma ().

Inappropriate Study Design (papers 8-9) Howard et al. used a bite plate to measure forces at the TMJ, which required firm closure of the mouth on the plate. Since jaw opening is integral to the mechanism of injury at the TMJ during whiplash (), having the subjects keep their mandible firmly elevated during the crash testing defeated the purpose of the study, and the results are meaningless with regard to the actual forces sustained at the TMJ during in vivo whiplash trauma.

Rosenbluth and Hicks gave no rationale for comparing whiplash trauma to rope skipping. The maximum acceleration reported in the x vector was 3.5 g for the seven year-old, and approximately 1 g for the 29 year-old, far less than ranges of acceleration reported by other authors for low speed rear-impact crash testing (6-14.5 g) (21,27). The difference between the acceleration noted for the child in comparison with the adult may be artifactual, since the helmets were secured to the subjects with a single strap under the chin, an arrangement that may have allowed for excessive movement between the helmet and the head (see figure 2).

Unsupported Conclusions (papers 9 and 10) Howard et al. (1995) compared the acceleration forces measured at the TMJ during a low velocity rear-impact collision to those of mastication, concluding that the non-injurious forces of mastication were far greater than those of whiplash trauma. However, the

authors did not study acceleration forces specifically at the TMJ, and thus cannot compare the forces measured in their study to those found with mastication, as mastication produces a differential acceleration between the cranium and the mandible. Since the jaw was closed in this study, the mandible was accelerated at the same rate as the cranium and no differential movement for the two osseous components of the TMJ was allowed. There was no scientific support for the conclusions of the authors regarding TMJ injury potential in the methods or results of this study.

Castro et al. noted symptoms of whiplash injury in 29% of their study subjects, yet ignored their study results when concluding that similar impacts were harmless. The authors contradicted their own study findings in their conclusions.

BIOMECHANICAL STUDIES

1. Allen et al. studied the acceleration forces of common movements in eight volunteers with triaxial accelerometers affixed to a helmet (). They reported peak accelerative forces that were measured while subjects "plopped in a chair" that were similar to accelerative forces recorded during published accounts of volunteer crash testing. Citing the results of their study, the authors stated that "no-damage accidents," like the common movements examined in the study, were unlikely to cause injury.

Methodologic Errors

Unsupported Conclusions Allen et al. concluded that whiplash trauma and ordinary daily movements were comparable, even though none of the movements studied duplicated the vector or force of whiplash trauma. The majority of acceleration in a rear-impact crash is in the x vector, that is, front to back. The largest single acceleration reported by Allen et al. was 10.1 g in a diagonal vector (54.9 degrees from horizontal) during "plopping in a chair." (See figure 3) However, the x vector component was only 5.6 g. In "Table 2" of Allen et al.'s paper, the mean x vector acceleration of plopping in a chair was 3.3 g, the highest mean x vector acceleration of all of the movements. In actuality, Allen et al. reported that 10 of the 13 movements studied had mean x vector accelerations less than 2 g. In comparison, West et al. reported a range of peak acceleration at the head during crash testing of six volunteers of 6 to 14.5 g (at nine km/h EBS) (27). Siegmund et al., in the largest published crash test to date, reported 6.7 to 12 g's of peak head acceleration among 39 subjects crash tested at eight km/h delta V (21). Additionally, the duration of peak acceleration of the movements studied by Allen et al. (approximately 1 millisecond) is not comparable to the duration of peak acceleration measured during whiplash trauma (70 milliseconds) (21). Taking into account both components of acceleration (magnitude and duration), whiplash trauma produces more than 150 times greater peak accelerative force than plopping in a chair.

Misleading Illustration In Allen et al.'s illustration of the acceleration forces measured while "plopping in a chair," the authors showed a human head apparently moving into extension, with an arrow traveling rearwards through the head, and "10.1G" labeled at the arrow head (see figure 3). However, the legend of the figure parenthetically states "the apparent axis of rotation of the head in this schematic is not the true motion of the head. It is an expression of the acceleration forces." In spite of the disclaimer in the legend, it appears that the authors are attempting to convince the reader that "plopping in a chair" produces the same vector and magnitude of acceleration, as well as movement at the head, as a rear-end collision.

Inappropriate Study Design Allen et al. did not give a rationale for comparing common movements that do not usually cause injury to whiplash trauma, which results in 2.9 million injuries annually. By its design, Allen et al.'s study could not yield any information about whiplash injuries, since neither whiplash injuries nor the mechanism of injury in whiplash injuries was studied.

2. In their paper on the theoretical biomechanics of temporomandibular joint during whiplash trauma, Howard et al. (1991) stated that "head accelerations produced by forces in the neck (extension-flexion motion) … will generate forces in the temporomandibular joints that… are of substantially lower magnitude than the forces encountered routinely with normal mastication ()." They also stated that the normal motion of chewing produced "greater potential to produce traumatic injury" than whiplash trauma.

Methodologic Error

Inappropriate Study Design In this paper, the authors theorized that extension of the head with the mouth closed would not cause injury to the TMJ. While this may be true, the most widely accepted and researched model of TMJ injury during whiplash centers around jaw opening during cervical extension, a motion that leaves the TMJ much more susceptible to posterior joint and intra-articular disc injury than when it is closed (35). The comparison that Howard et al. makes between the forces acting on the TMJ during whiplash trauma and the normal forces of mastication is fundamentally unsound. The position of the joint at the point of maximum force (closed) as well as the direction (cephalad) and type (compression) of the force during mastication cannot be meaningfully compared with the position of the joint (open) and the direction (posterior) and type (shear) of force during whiplash trauma to the TMJ.

Discussion

The methodologic flaws most frequently found in the reviewed studies were non-representative study sample (60% of studies), inadequate study size (60% of studies), non-representative crash conditions (50%), and inappropriate study design (45% of studies). Other flaws found were unsupported conclusions (25% of studies), unsubstantiated/unreferenced claims (15% of studies), misquoted literature (5% of studies), improper use of terminology (5% of studies), and misleading illustration (5% of studies) (see Table 2).

All of the papers that had non-representative study samples and crash conditions, inadequate sample size, and other errors resulting in poor internal validity (meaning that bias was present) also had poor external validity (lack of generalizability) as a result. In other words, if the study methods were significantly flawed, the results of the study could not be extrapolated to any population outside the study.

While the majority of studies that were reviewed for this critique were found to be lacking in study numbers, it is doubtful that any study size or design will define a threshold for whiplash injury, because it is probable that one does not exist. This presumption is based on the confirmed existence of numerous risk factors for whiplash injury that contribute to a highly variable individual susceptibility to injury.

Variables intrinsic to the injured occupant that have been identified as risk factors for injury presence, severity, and duration following whiplash trauma are female gender (,) increased age (), preexisting degenerative changes in the spine (), out of position occupant in the vehicle during impact (), rotation of the head during impact (), lack of preparation prior to impact (43,), and a slender physique (1), collectively referred to as Intrinsic Injury Risk Factors (IIRF) for the purposes of this literature review. Risk factors for injury extrinsic to the occupant are direction of impact (,), presence and position of a head restraint (1,), and presence of a shoulder restraint (,) , referred to as Extrinsic Injury Risk Factors (EIRF). Acceleration forces interact with the above mentioned risk factors, as well as Unconfirmed Probable Risk Factors (UPRF) such as car seat construction and bumper dynamics, to produce injury. The number of meaningful permutations of the IIRFs, EIRFs, and UPRFs is conceivably in the thousands or tens of thousands, making volunteer crash testing a highly unlikely study design for delineating an injury threshold for an entire population.

Conclusions

The results of the current literature review and critique suggest that the methodology employed by authors attempting to refute the validity of whiplash syndrome is flawed generally. With only a few exceptions, however, the studies reviewed contained other facets that employed relatively sound methods and that contributed to the knowledge base of whiplash injuries and biomechanics. Therefore, it is important to reiterate that the current critique only evaluated study methodology as it related to statements refuting whiplash syndrome.

It may be concluded, as a result of the current literature critique, that there is currently no epidemiologic or scientific basis for the following statements:

- acute whiplash injuries do not lead to chronic pain
- chronic pain resulting from whiplash injuries is usually psychogenic

- whiplash injuries are unlikely to result in chronic pain in countries where there is no compensation for injury
- rear impact collisions that do not result in vehicle damage are unlikely to cause injury
- whiplash trauma is biomechanically comparable to common movements of daily living
- there is insufficient force generated at the TMJ during whiplash trauma to cause injury
- TMJ injuries are not associated with whiplash trauma
- there is a direct relationship between vehicle damage and the probability of developing chronic pain following whiplash trauma
- chronic pain following acute whiplash injury is caused or worsened by treatment and diagnostic testing the risk of chronic neck pain among acutely injured whiplash victims is the same as the prevalence of chronic neck pain in the general population

As the body of whiplash literature increases, papers with findings that support one side or another of the legal debate over the validity of whiplash syndrome are increasingly likely to be used in legal settings. Editors and manuscript reviewers need to be alert for whiplash papers with flawed methodology, or that over-extrapolate their findings. The purpose of the present critique is to provide an overview of some of the weaknesses and the strengths of the whiplash literature..