What Every Attorney Should Know About Probability, Coincidence and Causation in Litigation

Michael D. Freeman, PhD, Forensic Research & Analysis, Portland, Oregon
Clinical Review: Focused

A Systematic Approach to Clinical Determinations of Causation in Symptomatic Spinal Disk Injury Following Motor Vehicle Crash Trauma

Michael D. Freeman, PhD, MPH, DC, Christopher J. Centeno, MD, Sean S. Kohles, PhD

Clinical determinations of causation in cases of intervertebral disk (IVD) injury after a motor vehicle crash (MVC) are often disputed in medicolegal settings. No published systematic guidelines exist for making such determinations, which has resulted in infringement by nonclinical personnel into injury causation evaluations, a traditionally clinical activity. The result is causal determinations that are potentially disconnected from clinical observations of injury. The purpose of this review was to evaluate the current literature on causation, causal determinations after trauma and IVD injury after MVC, and to develop a practicable, logical, and literature-based approach to causation determinations of symptomatic IVD injury after MVC. The results of the review indicate IVD injury can result from any MVC regardless of magnitude, thus meeting the first criteria of causation, biologic plausibility. Individual determinations of causation depend entirely on the temporal association between the collision and the symptom onset (the second criterion) and a lack of a more probable explanation for the symptoms (the third). When these causal elements are met, clinicians can assert causation on a “more probable than not” or “reasonable probability” basis. Because of a lack of an established or reliable relationship between collision force and the probability of IVD injury the investigation of collision parameters is not a useful adjunct to causal determinations.

INTRODUCTION

The origin of injury to the intervertebral disks (IVD) of the spine is a common source of dispute in medicolegal circumstances. Whereas treating clinicians typically make determinations of causal relationships between a trauma and an injury primarily based on the patient-related history of the traumatic event and the onset of symptoms, opposing experts frequently rely on population-based assumptions as a basis for disputing causal relationships. This scenario is observed most often when disk injuries and symptoms are attributed to motor vehicle crashes (MVC), because many crash-related injuries result from the negligence of another driver; thus, the nature and extent of an occupant’s injuries may be contested. There is typically a strong third-party (insurer) interest in minimizing compensatory payments on behalf of the individuals who are at fault for the crash. As a result, an adversarial system has developed in the United States and many other industrialized nations, with the injured patient and his or her treating physicians on one side and the party responsible for the crash, his or her insurer, and the insurer’s medical and scientific consultants on the other side [1,2]. Although some clinicians prefer to avoid causal determinations altogether, it can be argued that such determinations are part and parcel of regular medical care. To avoid such a determination solely because it may be later disputed by a party with interests opposing those of the patient is to avoid a duty to the patient; a duty that can have a significant impact on the patient’s well-being. This is not to say that dissemblance and fraud do not exist in the patient population; however, when there is no evidence of dishonesty, there is no reason to assume the presence of fraud simply because a patient is claiming injury after a MVC that was caused by another’s negligence (the most common setting for a legal claim for monetary damages). Such events are, for the most part, randomly distributed in the population that uses motor vehicles and not selective for people...
in whom fraudulent behavior is more common. When allegations of fraud are made in a medicolegal context, despite the absence of evidence of fraud, it is reasonable to question the motivation behind the assertion, because forensic consultants who provide partisan opinions are no more immune to secondary gain pressures than are litigants claiming injury [3].

There are some authors who maintain that for a determination of cause and effect to be made between a diagnosed injury and an MVC, a biomechanical analysis of the forces is necessary to determine whether or not a particular collision had the potential to cause the injuries [4] or that a detailed crash reconstruction is necessary before a causal association can be determined [5]. In contrast, most authors who have documented IVD injuries associated with an MVC make no mention of a biomechanical force analysis or a crash reconstruction [6,7]. The latter approach is most typical of the clinically pragmatic approach to causation. If it is known that an injury can be caused by a trauma, then there is no practical or logical purpose served in quantifying the degree of the trauma in order to assess the risk of injury after the event. Even injuries that are deemed highly unlikely or improbable are by definition still slightly likely or possible, and thus injury presence is typically evaluated after the fact clinically, and not called into question by a post-hoc suggestion of infrequency.

For the purposes of the present inquiry, a Medline search for the years 1980 through 2009 (as of July 2009) using the medical subject headings (MeSH) terms intervertebral disk, intervertebral disk displacement, and disk, herniated as well as the non-MeSH term cervical disk, in conjunction with “injury” and/or “trauma,” along with “cause,” “causal,” and “causation” in various combinations did not elicit any publications that addressed an organized approach to determinations of causation between disk injury and MVC exposure. It appears that, at the present time, there is no widely accepted scientific or clinical standard for such determinations.

The purpose of this article is to assess the evidence for the causal relationship between disk injury and MVC trauma based on the published literature on causal standards and the science of both disk injury and MVC trauma. Further, the authors propose a systematic and scientifically based approach to individual determinations of causation.

**PRIOR PUBLICATIONS ON CAUSATION**

Determinations of causal association between a noxious exposure (microorganism, chemical, trauma, or other) and a disease or injury outcome in populations is an area of considerable interest to epidemiologists and others who study such issues. Individual clinical observations of cause and effect can serve as a reasonable basis for a case study and suggestion of a relationship. It is, however, difficult to draw generalizable conclusions from such reports. Case series provide stronger evidence of causation; however, without a control group of unexposed subjects, there is no way to determine which effects can be isolated to a particular exposure. Even when case-control studies, in which diseased or injured subjects are compared with healthy subjects for level of exposure to a suspected noxious agent, show positive correlations between exposure and disease or injury outcome, the results cannot be interpreted as validated evidence of a causal relationship, as confounding factors may exist that can obscure true relationships. An example of confounding would be the relationship between coffee drinking and lung cancer. Coffee drinkers may have a higher rate of lung cancer than non-coffee drinkers, but only because smokers are more likely to drink coffee than nonsmokers, not because coffee is causally linked to lung cancer.

In an attempt to more clearly identify causal relationships between noxious exposures and outcomes in populations, Hill outlined a set of 9 criteria that needed to be considered to reasonably conclude that a particular noxious exposure could result in a particular disease [8]. Hill’s criteria have served as a kind of seminal gold standard, in that they have been adopted and modified by many others, including the Food and Drug Administration, academic investigators, and pharmaceutical companies, as a means of approaching issues of causation systematically [9]. Based in part on Hill’s work, Miller et al proposed a 4-stage process for identifying environmentally caused rheumatic disorders, an application that is more applicable to the goals of this article, because it allowed for assessments of causation in individuals rather than as a population-based construct [10]. McLean subsequently adapted Miller et al’s work for causation in instances of fibromyalgia after MVC [11]. All causal criteria can be distilled to a minimum of 3 common and essential elements, which are as follows [12,13].

1. There must be a biologically plausible or possible link between the exposure and the outcome. For example, trauma and fracture are plausibly linked, but trauma and leukemia are not. Plausibility is a low threshold that is exceeded with relatively weak evidence, such as from small observational studies (case studies or case series with small numbers of subjects) or from the results of well-designed experiments with many subjects. Biologic plausibility only pertains to whether an outcome can possibly result from an exposure, and is unrelated to the rate or frequency of the outcome. Thus, evidence of low incidence is not evidence of biologic implausibility.

2. There must be a temporal relationship between the exposure and the outcome. The outcome cannot preexist the exposure; however, the outcome of interest may have preexisted the exposure in a less severe form that was worsened by the exposure. Such a determination requires an accurate documentation of the signs and symptoms of the condition of interest both before and after the exposure of interest. Additionally, the outcome cannot postdate the exposure by a period that is considered, from a clinical perspective, to be too long or too short to relate the two. This determination is highly dependent on the specifics of any case. For example, an injury to a nerve root may cause immediate pain, but electromyogram changes may lag behind by weeks or months. The deter-
A Causation Algorithm for Disk Injury after MVC

In this section, the authors present a practical guideline for clinical determinations of causation in symptomatic disk injuries after an MVC, based on the application of the aforementioned 3 causal elements to the facts and findings in an individual case. The following definitions and assumptions are made. An IVD injury is defined by all of the following characteristics [14,15] (for this definition, “injury” is used only to define a disk that has become symptomatic after a discrete loading event of any magnitude, and the cause of the injury is not addressed).

1. It is symptomatic. This may include classic symptoms of segmentally appropriate radiculopathy, regional axial pain, or diskogenic referral of pain in a nondermatomal distribution.

2. There is clinical imaging evidence (magnetic resonance imaging [MRI], computed tomography, or diskography) that the symptomatic disk fits the generally accepted definition of a degenerative/traumatic disk [14]. This includes anular tears, herniations, and degeneration. Although there are findings specific to disk trauma, such as vertebral endplate fracture and edema presence in the narrow adjacent to the endplate, many traumatically initiated disk injuries are indistinguishable from degenerative changes in the disk [16]. Because of mediocre sensitivity and specificity (i.e., true- and false-positive rates), imaging findings alone are insufficient evidence for either a causal or noncausal determination in a case of suspected or disputed traumatic disk injury [17,18].

MVC trauma is defined as follows.

1. The patient was an occupant of a motor vehicle at the time of a collision or event.

2. The collision or event transmitted a force through the patient’s spine over a very short time, typically 70-120 ms [19]. Note that the magnitude of the force is not addressed in the definition, because any degree of forceful loading could potentially produce a symptomatic disk injury in a sufficiently susceptible individual.

There are 2 most likely scenarios in which a forcefully loaded IVD may become symptomatic.

1. A healthy disk with little or no changes associated with degeneration is subjected to a significant load resulting in disruption of the annulus and possibly instantaneous migration of nuclear material posteriorly toward the disk periphery and into the epidural space or the lateral recesses. This is the scenario most likely to produce injury to tissue surrounding the disk, including vertebral endplate fracture and ligamentous disruption, and is most likely to be associated with higher energy crashes with significant vehicle damage [20].

2. A degenerated disk (defined as a disk in which one or more of the following are present to some degree: desiccation, fibrosis, narrowing of the disk space, diffuse bulging of the anulus beyond the disk space, anular tears, endplate sclerosis, and osteophytes at the vertebral apophyses [14]) is loaded in a manner that initiates a symptomatic response. A significant proportion of the asymptomatic population has some degree of disk degeneration as seen on MRI; Matsumoto et al have performed the largest study to date on the topic, with MRI scans of the cervical spines of 497 asymptomatic subjects [21]. These authors describe a progressive direct relationship between the presence and severity of degenerative changes and age, with the disk degeneration noted in a range of 17% men and 12% of women in their 20s to 86% and 89% of men and women older than 60 years. Other authors have described similar findings with broad ranges of values depending on the age of the subjects; from virtually no disk degeneration in subjects <30 years of age, to 14% in those 30-40, and 62% in subjects >40 years of age [22,23]. Similar observations have been published with regard to lumbar spine degeneration among asymptomatic patients. Jensen et al described a cohort of 98 subjects who underwent lumbar MRI, noting that only 36% of subjects had no abnormal disks. Fifty-two percent of the scans were interpreted as having a bulge at a minimum of one level, 38% had abnormality at more than one level, 27% had a protrusion, and 1% had an extrusion [24]. These findings are consistent with those reported by other authors [25].

Causal Elements

Biologic Plausibility. Can an MVC of any severity cause a disk injury? Crash testing of intact cadavers at accelerations recorded for no-damage rear-impact collisions has demonstrated IVD injury that cannot be detected by conventional
imaging such as CT and MRI [26]. Although the biomechanical literature on experimental loading of disks in cadaveric sled testing has demonstrated disk injuries even in relatively low accelerations (3.3 and 4.5 times the force of gravity or ‘g’ [26]), such ex vivo testing of healthy disks does little to demonstrate minimal thresholds for injury in live human populations exposed to real-world crashes. No cadaveric testing can duplicate the significant variation in the manner in which loads are transmitted to the human spine in a MVC, given the variation in occupant position, restraint systems, vehicle interiors, and muscle loads. Additionally, the potential for variation in the condition of an individual IVD, including the ability of the disk to resist forceful loading from an external source, is too large to ever define in an experimental setting.

A relevant literature search revealed a complete absence of any publications suggesting that symptomatic IVD derangement cannot result from MVC exposure of any magnitude. To the contrary, injury to the disks of the cervical and lumbar spine has been documented in the relatively low level accelerations associated with little to no-damage collisions, roller coaster rides, and even sneezing [27-30]. It is reasonable to conclude, as a general precept, that the forceful loading of the spine that can occur in any MVC is a biologically plausible (possible) cause of symptomatic disk injury.

**Temporal Association.** Temporality is the strongest evidence of causation in evaluating the patient with post-MVC disk injury. It must be first established that the MVC preceded the onset of symptoms attributed to the disk injury. The exception to this rule is when a previously symptomatic disk is exacerbated by an MVC to the point that the course of care is significantly altered (ie, a previously diagnosed non-surgical disk condition becomes surgical directly after an MVC). In such cases, the determination must be made clinically as to whether the disk symptoms were likely to have worsened to the point that the surgical intervention was inevitable absent the MVC. The mere fact that the disk was previously symptomatic is not sufficient to draw the conclusion that it would have required surgery, given the good outcomes for patients with diskogenic symptoms who use a variety of surgical and nonsurgical treatments [31-33]. The symptom onset must be in reasonable proximity to the time of the MVC; however, disk injury symptoms may initially present as identical to symptoms of spinal strain or sprain, and it may be weeks or months before an MRI is performed and the disk injury diagnosis is first seen [34]. Additionally, disk injury symptoms may be progressive, as an injury to the annulus may progressively allow nuclear migration toward neurologic structures over a period of time while the patient is in weight bearing and active. Determinations as to whether delayed symptoms fulfill the temporal association criteria must be made by clinicians on a case-by-case basis.

An obvious weakness of the temporality criterion is that it can only be established from the history given by the patient, and patient histories can be erroneous or falsified. The former is unlikely in the case of injury, however, because one of the hallmarks of injury is that there is a close temporal relationship between cause and effect (this lack of clear temporal proximity in repetitive traumatic exposures is the reason that injuries such as carpal tunnel syndrome are classified as diseases) [35]. This close temporal relationship means that it is unlikely that a patient will misattribute symptoms to a traumatic cause. As a practical matter, the most reliable indicator of when a patient began to have symptoms after a trauma is the patient. This is not to say that patients do not get details concerning an injury event wrong; patient perception or recollection of vehicle speeds, crash sequence, and other details of a collision may be inconsistent with the facts of a case for reasons other than untruthfulness. For example, an occupant who is injured in a minimal damage rear end collision may believe that the impact must have been at high speed despite a lack of physical evidence that this was the case.

With regard to the truthfulness of the patient, it is not the role of the clinician to investigate the veracity of the history given by the patient. Absent some ancillary indication of deception, clinicians are generally safe to believe their patients. The rate of fraud of any kind in auto insurance claims for treatment benefits (personal injury protection coverage) was estimated to be 6% for 2007 [36]. Even if every instance of fraud in auto insurance cases consisted of a patient giving a false history, clinicians would still be justified in believing their patients 94% of the time.

Intuitively, it makes sense that the closer the onset of symptoms is to the time of the collision the stronger the causal relationship between the MVC and the disk injury, and the strength of this relationship can be quantified using an indirect approach. For example, in a case of a 40-year-old male with neck pain and cervical radiculopathy that arose within 12 hours of a rear impact collision (later attributed to MRI findings of disk herniation), assuming no intervening trauma, it can be postulated, as a tautology, that either the crash caused the disk injury or that it was coincidental to the collision, regardless of cause. Thus, the probability that the crash caused the disk injuries [P(disk\_MVC)] plus the probability the disk symptoms are coincidental to the crash [P(disk\_COINC)] when added together account for all of the possible causes (100% or 1) of the disk injury and associated symptoms. If [P(disk\_MVC) + P(disk\_COINC) = 1] then rearranging the terms gives [1 − P(disk\_COINC) = P(disk\_MVC)], meaning that if the probability that a disk injury occurred coincidentally on the same day of the MVC can be determined then the probability the MVC caused the disk injury can also be determined indirectly.

The probability of 2 unrelated events occurring in close temporal proximity can be calculated by multiplying the probability or odds of one times the other (odds are the ratio of 2 probabilities). For example, the odds of rolling 2 sixes in a row with a fair die are [(1:6) × (1:6) = 1:36]. This is because there are 36 possible combinations for any 2 rolls, including 1 and 1, 1 and 2, 1 and 3, etc, and only one of the combinations is a 6 and 6.
In a similar fashion the odds of a disk injury occurring coincidentally on the day of a collision can be calculated for the example given earlier. The annualized odds of involvement in a MVC for a 40-year-old male are 1:14 (once every 14 years), and the daily odds of crash involvement are approximately 1:5100 (one crash every 5100 days) [37]. The odds of the spontaneous development of disk symptoms are a bit more difficult to estimate. For example, it can be said that the patient had a single-day episode of neck and arm pain 12 months before the MVC, with no intervening episodes until the time of the collision. Thus, the odds that the symptoms would recur on any single day were no greater than 1:365. Multiplied together these 2 odds yield an odds of a disk injury occurring coincidentally on the day of the collision of 1:1,861,500. Conversely, the odds in favor of a causal relationship between the collision and the disk injury, given the onset of symptoms on the day of the collision, are 1,861,500:1.

Lack of Likely Alternative Explanations

Alternative explanations for a disk injury include an intervening trauma that followed the MVC and preceded the symptoms indicative of disk injury. Another alternative explanation is the insidious onset of symptoms, possibly associated with a trivial perturbation such as sneezing. Although minor forces are possible causes of disk injury, they cannot be preferentially selected over the substantially greater forces of a collision when the two have occurred in close temporal proximity. If, for example, a patient sustains what is diagnosed as a lumbar strain in a collision and a week later sneezes and has a sudden onset of radicular pain associated with a disk extrusion, then it is most likely that the disk was injured in the MVC and the sneeze only prompted a progression of symptoms. It is not reasonable to point to minor forces associated with daily activities as a likely cause of a symptomatic disk in preference to the significantly greater trauma associated even with minimal and no-damage MVCs that is temporally relevant to the symptoms, particularly when the patient performed such minor activities with no difficulty prior to the MVC. A “likely” alternative explanation is one that is most probably causal and the best explanation to fit all of the facts at hand, not one that is only a possible cause with no historical or temporal evidence to link it to the symptomatic disk injury.

DISCUSSION

Individual determinations of causation for IVD injury after MVC by clinicians require only that temporal association and lack of likely alternative explanation have been met for a particular symptomatic disk injury. It must be noted that the model of causation presented herein suffers from the fact that there is no other alternative against which it can be evaluated for accuracy. Clinicians receive little or no formal education in medical school or residency regarding a systematic approach to causation. Thus, the authors recommend that guidelines presented herein be evaluated, modified if necessary, and considered for adoption by consensus by appropriate scientific and medical organizations.

CONCLUSIONS

Individual determinations of causation for disk injury after MVC are most appropriately conducted by clinicians, based primarily on an evaluation of the temporal association be-
between the MVC and the symptom onset. The lack of a valid or meaningful collision force threshold below which it can be said that a disk injury will not occur means that investigation of collision parameters for the purposes of injury causation assessment is a pointless endeavor.

REFERENCES

An evaluation of applied biomechanics as an adjunct to systematic specific causation in forensic medicine

Michael D. Freeman¹ and Sean S. Kohles²

¹Department of Public Health and Preventive Medicine, Oregon Health and Science University, Portland, Oregon, USA, and Institute of Forensic Medicine, Faculty of Health Sciences, Aarhus University, Aarhus, Denmark

²Regenerative Bioengineering Laboratory, Department of Mechanical and Materials Engineering, Portland State University, Portland, Oregon, USA

Received September 25, 2010, accepted (after revision) December 14, 2010, published online July 29, 2011

Summary. Biomechanical tests of post hoc probability have been proposed by prior authors as reliable tests of causation in forensic settings. Biomechanical assessment of injury kinetics and kinematics is a potentially important tool in forensic medicine, but there is also the potential for misapplication. The most reliable application is when biomechanical analysis is used to explain injury mechanisms, such as how an injury may have occurred. When a biomechanical analysis is used as a means of determining whether, rather than how an injury has resulted from a traumatic exposure, then a lack of reliability of the methodology limits its application in forensic medicine. Herein, we describe a systematic assessment of causation by adapting established general causation principles to specific causation scenarios, and how biomechanical analysis of injury mechanics is properly used to augment such an approach in conjunction with the principles of forensic epidemiology. An example calculation of relative risk associated with cervical spine injury is provided as a representative probabilistic metric for assessing causation. The statistical benefits and limitations of biomechanical analysis are discussed as an adjunct to forensic medicine.

Key words: Forensic medicine, biomechanics, Hill criteria, causation, forensic epidemiology

Introduction

The use of biomechanical analysis of injury mechanisms as an adjunct to forensic medicine has been increasing over the past 20 years. Biomechanics may be simply defined as the study of the effect of mechanical energy on biological tissue. However, in the context of forensic medicine the discipline is used most often to define injury thresholds, and to match injury mechanisms with expected or observed injuries as a means of causal determination [1]. As an applied mathematics discipline, various categorical approaches define the fundamental assumptions typically applied in forensics such as rigid body kinetics/kinematics or deformable body deformations (Tab. 1). These initial assumptions guide the appropriate mathematical treatment of the case at hand. Biomechanics has been demonstrated as an important but limited role in forensic investigation of death and injury. The discipline can be very helpful and relevant for analyzing how injuries occur and in differentiating between competing injury mechanisms for observed injuries, but it can result in error-prone conclusions for other applications [2].

We have previously described how forensic biomechanics is most reliable when used to explain how observed injuries occur, less reliable when evaluating competing hypothetical explanations for an observed injury, and least reliable when used to refute the presence of observed injuries [2]. As an example, a group of authors have described 20 femur fractures occurring in restrained occupants in frontal collisions, some at unexpectedly low speeds (16 km/h) [3]. An analysis of the collision forces was compared to prior biomechanical experimental ex vivo data that predicted a low or non-existent risk of fracture. The
authors concluded that pre-impact muscle forces were the best explanation for the occurrence of the fractures. In the context of a forensic setting, a biomechanical analysis using the same data might be used as a test of whether an occupant was telling the truth about using a seatbelt in a 16 km/h frontal collision in which there had been a femur fracture. This would constitute a hypothetical application of biomechanics, in which the probability of injury was assessed for a hypothetical scenario (seatbelt use versus nonuse). The result of the forensic analysis, based on the results of the cadaveric studies cited by these authors, would be that there was a 0% probability of femur fracture if a seatbelt had been used, and thus the seemingly valid, but ultimately incorrect conclusion that the occupant was not telling the truth about the use of a seatbelt. The same analysis might also be used, less logically, to refute the presence of the injury, with a conclusion that a claimed femur fracture could not have occurred if a seatbelt was worn because of the 0% experimentally derived probability of fracture with a restraint, and thus if the seatbelt was worn the fracture did not occur.

The example demonstrates the difficulty with the application of forensic biomechanics; although the discipline is often used for making causation determinations, the validity of the conclusions is highly dependent upon the circumstances in which they are applied. Much of this difficulty stems from the fact that the purpose of a forensic biomechanical analysis is often to estimate injury risk, and risk has a highly circumscribed role in evaluating causation.

An important consideration with a forensic biomechanical analysis of causation is how the results correlate with a forensic epidemiologic analysis of causation. Epidemiology is generally defined as the study of disease and injury in populations, and is the scientific basis of general or population-based causation [4]. The field of Forensic Epidemiology applies epidemiologic (population-based) data as a basis for evaluating the consistency of the findings in an individual case with what is plausibly associated with a particular injury mechanism [5]. Forensic Epidemiology addresses the science of specific or individual causation.

The purpose of this paper is to assess and describe how forensic biomechanics, when used to assess the causal correlation between a potential injury mechanism and an observed injury, fits within the framework of a systematic (epidemiologic) approach to general and specific causation.

Material and methods

Background and historical context
Causal assessments are typically performed in two major realms. The first is an epidemiologic based assessment of causation in populations (general causation). Questions such as “does cigarette smoking cause lung cancer” are answered by carefully designed

<table>
<thead>
<tr>
<th>Category</th>
<th>Sub-category</th>
<th>Metric</th>
<th>Forensic application</th>
<th>Representative publication</th>
</tr>
</thead>
<tbody>
<tr>
<td>Rigid body</td>
<td>Kinetic</td>
<td>Impact forces</td>
<td>Blunt trauma; trip/fall</td>
<td>Whittle et al., 2008 [35]</td>
</tr>
<tr>
<td>Deformable bodies</td>
<td>Material failure (tissue, engineered)</td>
<td>Deformation; failure mode</td>
<td>Motor vehicle crash; bone/joint fracture</td>
<td>Hayes et al., 2007 [18]</td>
</tr>
<tr>
<td></td>
<td>Material properties (elastic, viscoelastic)</td>
<td>Moduli; rate of loading</td>
<td>Impact rates</td>
<td>Provenzano et al., 2002 [38]</td>
</tr>
<tr>
<td></td>
<td>Composite materials</td>
<td>Homogeneous/anisotropic geometric arrangement</td>
<td>Hard/soft tissues failures; automotive manufacture</td>
<td>Roy et al., 2004 [39]</td>
</tr>
</tbody>
</table>
epidemiologic studies showing repeated and consistent relationships between suspected cause and observed disease or injury effect.

The second type of causal assessment is the evaluation of cause and effect in individuals (specific causation), most often performed by clinicians in a clinical setting, but also a frequent issue of contention in forensic settings. The two are inextricably interwoven, inasmuch as specific causation depends on principles of general causation, and general causation is based upon a foundation of individual cases of specific causation. Clinicians make determinations of causation that are centered around diagnosis and history, and based upon interaction with patients. Epidemiologists typically assess specific causation in a post-clinical setting as a means of determining whether an individual case fits the definition for inclusion in a larger group of similar patients or subjects. Biomechanical analysis is most often used in such a setting after a causal determination has been made, as a means of explaining injuries that have been observed medically and catalogued epidemiologically.

Forensic Epidemiology has historically served as a basis for quantifying causal determinations in legal and forensic settings. General and specific causation approaches employing epidemiologic data and methods have been applied in legal rulings by U.S. courts including cases in administrative law, civil law, criminal law, and public health law [6]. As an example of the admission of epidemiology as relevant evidence, courts have accepted as evidence reports of epidemiological studies conducted by the U.S. Centers for Disease Control that analyzed the association between tampon use and the risk of toxic shock syndrome as a means of specific causation determination [7]. In reaching its decision, the Court relied on several prior cases that had permitted the admission of epidemiological evidence for causal determinations. In another case, epidemiologic testimony was used to establish the time period during which an individual was at increased risk of contracting Guillain Barre syndrome following Swine Flu vaccination, a critical factor in determining both general and specific causation [8]. Epidemiologic evidence has also been used to address issues of causation pertaining to communicable disease transmission [9, 10]. Epidemiologic causation has been used in many other legal circumstances, ranging from the criminal context [11] to disputes over regulatory issues regarding exposure to toxic substances [12].

As an adjunct to causation, Forensic Biomechanics has a briefer history in the courts. Biomechanics has been applied in forensic settings primarily as an adjunct analytical approach in motor vehicle crash cases [13]. These forensic applications of biomechanical analysis have not been directed at absolute causation (i.e. whether injury occurred), but rather at assessing the probability of competing theories of injury. In 2000, Walz and Muser introduced the theory that a biomechanical analysis was an essential part of causal determinations of so-called “whiplash” injury after traffic crashes [14]. These authors suggested that without such an analysis a clinical determination of causation was not sufficient for a court setting. Multiple bioengineering approaches have been proposed for the forensic setting as a means of introducing a technology-based approach to causal analysis. These approaches include detailed anatomic distinctions in tissue-joint properties [15], three-dimensional computational analyses [16, 17], and tissue-joint failure threshold metrics [18]. The application of such methods may carry with them the implication of greater validity in a forensic setting because of the lay misperception that a high-technology approach to causation is in some way superior to the more traditional but low-technology clinical approach.

General causation
A working definition of causation represents a specific event as an antecedent event, condition, or characteristic that was necessary for the occurrence of the disease, injury, etc. at the moment it occurred, given that other conditions are fixed [19]. In other words, the cause of a disease or injury event is an event, condition, or characteristic that preceded the disease event and without which the disease or injury would not have occurred at all or would not have occurred until some later time. The scientific basis for general and specific determinations of cause and effect was introduced through the inductive canons of John Stuart Mill [20] and the rules proposed by the philosopher David Hume [21].

In the current era, a practical approach to causation was laid out in a systematic fashion by Sir Austin Bradford-Hill in 1965 [22]. Hill outlined nine criteria by which population-based determinations of causation could be made when there is substantial epidemiologic evidence linking a disease or injury with an exposure, e.g. smoking and lung cancer [23]. Hill’s criteria have served as the seminal basis from which virtually all subsequent systematic approaches to general and specific causation have been derived, including those for a variety of injuries including traumatic brain injury [24], carpal tunnel syndrome [25], needle stick injuries [26] and spinal disk injuries [27], inter alia.
Hill’s original nine criteria, and how they pertain to both general and specific causation, are as follows.

**Strength of association**
Hill considered strength of association to be the most important determinant of causation. Most simply stated, a strong association is more likely to indicate a causal relationship than is a modest or weak association. Strength of association can be measured in general causation by the percentage decrease of an illness or injury in society or in a specific population if the injury cause were to be eliminated [19]. This is also known as the etiologic fraction that the particular cause contributes to the total societal burden of the disease.

In specific causation, strength of association is determined by a comparison between the injury or disease risk of the exposure and the risk of alternative or competing explanations for the injury or disease given the temporal relationship between exposure and outcome. An example would be a death occurring 1 hour following the administration of an intravenous drug. The death may be very rarely associated with the drug (e.g. 1 death per 100,000 doses), and thus the strength of association, relative to general causation, is very low (only a very small proportion of total deaths would be avoided if the drug became unavailable). However, if the probability that other causes of death that would have been likely to act coincidentally between the time of the drug administration and the death (e.g. infection, sudden cardiac death, etc.) is even lower, then the relative strength of association of the drug to the death, in comparison to other causes, is very high.

The strength of association is measured by relative risk (RR), which is the ratio between the rate of injury or disease outcomes in the exposed population versus the rate of injury or disease outcomes in the unexposed population. In specific causation, RR is measured by the risk of the condition resulting from the exposure versus the risk of the condition arising at the same time as the exposure from some cause other than the prime causal suspect. In order to meet the “more probable than not” (>50% probable) criterion for presentation in a forensic setting, an exposure must have an RR of greater than 2.0, given the specific circumstances of the case in question (mathematically, a relative risk of 2.0 is the same as a 50% probability). A common example of the use of RR in a forensic setting is seen in the evaluation of causation in cases of disease that are associated with exposure to a toxin, such as cigarette smoking. For example, passive exposure to cigarette smoking has a relative risk for lung cancer of 3.4 for certain populations [28]. This means that among equivalent populations, there will be 3.4 cases of lung cancer among the exposed for every 1.0 case of lung cancer among the unexposed. Thus, among the exposed, 71% of the cases of disease (2.4 of the 3.4 cases) resulted solely from the exposure, and 29% of the cases (1.0 of the 3.4) occurred regardless of the exposure to secondhand smoke. Based on these data it could be concluded, in an individual case of lung cancer associated with passive exposure to cigarette smoke, that more often than not (>50% of the time) the lung cancer was caused by the smoking.

**Consistency**
In general causation, the repetitive observation of a causal relationship in different circumstances strengthens the causal inference. For example, a causal relationship between cigarette smoking and lung cancer is observable for all brands and varieties of cigarettes. Consistency is present in specific causation if other individuals have been observed with the same outcome following substantially similar exposures, for example in observational (epidemiologic) study.

**Specificity**
In general causation, specificity refers to the degree to which a factor is associated with a particular outcome or population. In his original paper, Hill famously referred to scrotal cancer in chimney sweeps as an example of specificity, as the condition rarely occurred outside that particular population. For example, in an intravenous drug death, the drug may have a low specificity for death, as a very small proportion of deaths are caused by overdose of the drug, and the cause of death is typically multifactorial. In contrast, a decedent found with a 3 cm circular depression in his skull and a hammer nearby has an injury that is highly specific for having been struck in the head with the hammer. Thus, specificity has value in specific causation when it’s present, but a lack of specificity does not imply lack of causation.

**Temporality**
Temporality is the sine qua non of injury causation, in that the exposure must precede the injury. Hill only refers to temporality as the basis for making certain that the “horse comes before the cart” in general causation. In specific causation, however, an additional parameter
of temporality is considered, and that is the latency between the exposure and the first evidence of the injury [29]. Evidence of injury presence must not follow the exposure by a time period that is considered too great to link the two. Further, the time between the exposure and the first evidence of the injury outcome serves as a basis from which to assess the incidence of competing coincidental causes, as described under Strength of Association.

Biological gradient
In general causation, this refers to the observation that the injury or disease outcome increases monotonically with increasing dose of exposure. This criterion is particularly applicable to toxic exposures and at middle levels of traumatic exposure, but at the extremes of exposure it is not particularly helpful for assessing causation. For example, from a general causation perspective, a fall from 5 m will result in fracture more often than a fall from 2 m, however there is unlikely to be an appreciable difference in fracture rate between a 200 and 300 m fall.

For specific causation, the criterion of biologic gradient has more limited practical applicability. While it can be concluded, as a general principle, that the incidence of injury will generally increase as exposure intensity increases (to a point), injury incidence only serves as the numerator of the relative risk assessment critical to an evaluation of Strength of Association, and only has meaning or relevance when compared with magnitude of the potential competing causes in the denominator.

Plausibility
For both general and specific causation, plausibility refers to the degree to which the observed association can be plausibly explained by known scientific principles. Hill did not put much weight in plausibility, noting that a hypothesized disease cause that is thought to be implausible today may be discovered to be plausible at some time in the future as a result of new scientific inquiry. A more accurate way to characterize the practical application of the plausibility criterion is that it is met when there is a lack of established implausibility (impossibility). For example, a brain tumor discovered the day following a head trauma is implausibly related to the trauma due to the nature of the disease and injury mechanisms, respectively. A common error in causation is to consider a rare outcome to be the same as an implausible outcome. Rarity is not the same as implausibility, since a particularly rare outcome may still result in a large RR favoring the prime causal suspect if the denominator of competing causes is substantially smaller than the incidence numerator.

If a causal correlation is determined solely based upon the temporal association between an exposure and outcome, but implausibility is well established, then the post hoc ergo propter hoc fallacy has been committed. This fallacy is commonly explained using the example of the rooster who crows and sees the sun come up and subsequently infers that his crowing was the cause of the sunrise. Because the association is implausible, the rooster’s inference is fallacious.

Just as important, however is the recognition of the false assertion of the post hoc fallacy, in which a rare outcome (low probability) is mistaken for implausibility (0% probability), or implausibility is incorrectly asserted, and causation is improperly rejected in the presence of a close temporal relationship between exposure and outcome.

Coherence
For both general and specific causation, a causal conclusion should not fundamentally contradict present substantive knowledge – it should “make sense” given current knowledge. In some ways this criterion is much like plausibility. To use the earlier example, a blow to the head will not cause a tumor to develop overnight, as this is so far beyond of what is known about the pathophysiology of tumors that it is not a coherent explanation of cause and effect.

Experiment
In some cases there may be evidence from randomized experiments on animals or humans, in which an exposure is removed and there is a corresponding change in the frequency of the outcome. Experimental evidence for causation is treated identically in both general and specific causation; when it is present it is helpful, but the absence of experimental evidence is not evidence against a causal relationship.

Analogy
In both general and specific causations, an analogous exposure and outcome may be translatable to the circumstances of a previously unexplored causal investigation. Hill noted, as an example of analogy, that the birth defects that have been investigated and found to be causally associated with thalidomide or rubella exposure make it easier to accept a cause and effect relationship between another drug or virus for which there was less evidence. Evidence of analogy can serve as evidence in favor of plausibility.
Errors of causation
Two errors that can result from a faulty specific causation analysis are the acceptance of causation when it is not present (a Type I or Alpha error), and the rejection of causation when it is present (Type II or Beta error). A Type I error can result from the miscalculation of relative risk, resulting in a failure to account for a more probable alternative explanation, or from the lack of acknowledgement of well established implausibility between exposure and outcome. A Type II error results from either the lack of consideration, or inaccuracy of the incidence of alternative explanations (the denominator of the relative risk ratio described below). Further, an erroneous conclusion of implausibility will also result in a Type II error.

Another potential source of error stems from the fact that both forensic biomechanical and forensic epidemiologic analysis of specific causation are, in most instances, dependent upon the accuracy of the medically determined diagnosis as well as the accuracy of the history of the traumatic exposure and onset of signs indicative of the injury outcome. In a forensic setting, unless there is evidence of substantive inaccuracy in clinical or historical records it cannot be assumed that it is present.

Results and discussion
Applied forensic biomechanics and systematic specific causation
As a practical matter, for determinations of specific causation Hill’s criteria can be simplified and grouped into 2 major categories: criteria that answer the question “could the exposure have caused the disease or injury outcome in this case?” (Biologic Plausibility) and those that answer the question “did the exposure cause the disease or injury outcome in this case?” (Causal Association Strength).

Biologic Plausibility – This collective criterion is meant to assess whether or not the exposure could reasonably have caused the disease or injury outcome (regardless of how often), and is a composite of the previously described plausibility, coherence, specificity, consistency, biologic gradient, analogy, and experiment criteria. No single factor must be present to establish biologic plausibility, but it implausibility cannot be a well-established fact.

In injury causation, a forensic biomechanical analysis may be useful in evaluating biologic plausibility. A description of the magnitude, direction, material properties, and rate of application of a load source can help explain how a particular injury may have occurred. Caution must be exercised with such an analysis, however, as it is primarily unidirectional. Plausibility refers to a probability of more than 0%, and implausibility refers to a probability of no greater than 0%. A lack of plausibility is neither, and thus implausibility (probability = 0%) and lack of plausibility are not equivalent. From a probabilistic perspective, however, a relationship for which plausibility is unknown is more likely to be plausible (>0%) than implausible (0%).

Hayes et al. have described a metric that they have called “biomechanical plausibility” which is intended to replace Hill’s plausibility criterion [18]. These authors describe biomechanical plausibility as the ability to demonstrate, using experimental data that forces from a possible injury event are sufficient to exceed the injury tolerance such that the injury can be said to have been “caused” by the event. Conversely, these authors interpret the inability to demonstrate biomechanical plausibility as an indication that the injury event could not have caused the injury in question. In drawing this conclusion these authors commit the “fallacy of the transposed conditional” by concluding that the absence of evidence of biomechanical plausibility of causation is equivalent to evidence of implausibility. This conclusion is incorrect and is in direct contradiction of Hill’s conclusion that he was “convinced that [plausibility] is a feature that we cannot demand [in causation].”

Implausibility is only established when cause and effect is ruled out because of the violation of a fundamental biologic principle, such as the universally accepted and incontrovertible maxim that head trauma cannot cause brain tumors to spontaneously appear overnight. A lack of experimental biomechanical data supporting a plausible relationship between a loading event and an injury does not equate to evidence of implausibility, but rather stands as a lack of biomechanical evidence of plausibility. This is not an obstacle to a causal analysis, particularly if plausibility is bolstered by the presence of coherence and analogy.

There are several reasons why the inability to demonstrate biomechanical plausibility does not equate to implausibility. Inherent in a determination of implausibility is a point estimate of 0% probability. The presence of bias and random error in experimental biomechanical data limits, to a certain extent, the extrapolability of the study results to circumstances and subjects like those in the experiment. Experimental biomechanical studies of animal, cadaver, and human
volunteer subjects produce results that describe only a part of the spectrum of injury response to a nearly infinite range and combination of injury scenarios in the real world (Fig. 1). Thus, a point estimate of 0% injury probability derived from experimental data is likely to represent a point in the probability spectrum of the real world response to traumatic loading that is greater than 0. Thus, implausibility in the context of causation is unlikely to be validly derived from experimental biomechanical data.

The caution advised in the preceding statement largely stems from the effect of bias and scatter on the extrapolation of tolerance specifications based in experimental data to real world occurrences as exclusionary boundary conditions. Most simply this means that experimental studies cannot contain the quantity and types of subjects and conditions that sufficiently represent the range of variety seen in real world subjects and conditions. For this reason, experimental biomechanical data can only serve as a basis for a determination of plausibility, but not implausibility.

**Causal Association Strength** – This criterion is based on Hill’s temporality and strength of association criteria. Temporality is defined by the facts in a particular case (most often gathered in a clinical setting), and cannot be augmented or defined by a technological analysis, although the independent taking of a history of the exposure and the onset of symptoms indicative of the injury in question outside of the clinical context may be useful for clarification. Once the determination has been made that the injury outcome followed the loading event exposure in time, the latency between the first indication of injury presence and the loading event must be quantified. The strength of association between the injury event and the injury outcome, expressed as a relative risk ratio, can then be assessed.

**Relative Risk (RR) calculation**

**RR numerator**
A biomechanical evaluation of potential injury forces can be used to estimate the absolute injury risk of an event, and this value serves as the numerator of the relative risk ratio. The methods used in a biomechanical analysis of loading of the human body depends on the manner in which the load was introduced to the body, as well as the nature of the reference data that are used for comparison to tolerance thresholds or from which injury risk is derived, depending upon the forensic application. Various types of biomechanical approaches may be categorized in their role of assessing absolute injury risk (Tab. 1).

The following scenario illustrates a hypothetical forensic biomechanical analysis resulting in an absolute risk estimate for the numerator of a relative risk calculation. Here, an occupant is diagnosed with a concussion following a frontal collision. The biomechanical analysis would begin with an estimation of the collision forces via reconstruction of the crash dynamics. This is accomplished by the application of equations of momentum, energy, and restitution (MER) as derived from Newton’s Laws [30]. The approach incorporates known (or assumed) variables of vehicle mass, initial vehicle velocities, and vehicle elasticity and deformation as a means of estimating vehicle speed changes and the potential effect (kinematics) on vehicle occupants. An injury metric, based upon observational data, can then be inferred from the resulting analysis of occupant-vehicle interactions. For the example, we can apply the Head Impact Power metric [31] to a head-dashboard strike, resulting in a HIPmax of 10 kW (Fig. 2).

At a HIP$_{max}$ of 10 kW there is a corresponding approximately 27% probability of concussion, based
upon observational data gathered from a cohort of professional football players exposed to head impacts [31]. This point estimate cannot be directly extrapolated to the general population because the underlying data are subject to the same limitations (bias and random error) that prevent the use of biomechanical data as a basis for a determination of implausibility. A means of compensating for the effect of random error is to calculate a confidence interval for the point estimate, which accounts for the size of the sample that serves as a basis for the point estimate. Applying a confidence interval to the data represented in Fig. 2 results in an injury probability range of approximately 9 to 58% for a HIP\text{max} of 10 kW (see arrows in Fig. 3).

As a side note, the injury investigation technique described by Hayes et al. asserts a threshold of 50% risk as a requirement for "biomechanical plausibility" as an ersatz form of Hill's plausibility criterion [18]. This approach is doubly incorrect: first, plausibility, by definition, is met whenever there is more than 0% probability of injury. It is fallacious to imply that an injury risk of less than 50% constitutes implausibility, as this is the equivalent of asserting that any value less than 50% is the same as 0%. Secondly, the method is incomplete, as it only addresses absolute risk in reaching a causation conclusion, and ignores the denominator of competing explanations required for a relative risk assessment.

As in illustrative example of the impracticality of such an approach, one can imagine a low speed traffic crash after which a driver is found to be deceased. Because of the nature of the collision, the biomechanical risk of the fatal injury is thought to be less than 0.01%. The method advocated by Hayes et al. would then force the erroneous conclusions that either the death did not occur as a result of the collision or that the determination that the occupant had died was erroneous. The example demonstrates the logical appeal of the RR approach advocated by Hill; if the competing risk of death during the same time period as the collision was less than the risk presented by the collision then the RR would favor the crash as the cause of the death, despite the low absolute risk.

**RR denominator**

The denominator of the relative risk calculation is the probability that the outcome would have occurred during the same time frame as the suspected exposure, had the exposure not occurred. Another way to think of the RR denominator is that it is the probability of an alternative explanation for the outcome. Occasionally, there may be a competing injury mechanism explanation that can be analyzed with a biomechanical assessment of forces. More often the only other explanation for an injury outcome aside from a suspected exposure is the coincidental (relative to the exposure) onset of the condition, an epidemiologic parameter. For the present application, coincidence is defined as the temporal or spatial alignment of 2 or more unrelated events. The probability of 2 or more unrelated events occurring at the same time, like the probability that two die rolled will yield a pair of sixes, is calculated by multiplying the probabilities by each other, based upon the 3rd Rule of Probability that the joint probability of two or more independent probabilities is the product of the probabilities [32]. Thus the probability that a single role of two die will result in a pair of sixes is (1/6 × 1/6 = 1/36 = 2.8%). In the same fashion, the joint probability of an injury coincidentally occurring at the same time as a traumatic exposure is the product of the probabilities describing each of the events.

**Calculating causation using relative risk and random match probability**

The specific causation approach described herein allows us to develop a mathematical definition of RR in
order to answer the question “did exposure A cause condition B?” by comparing all of the known and potential causes of B given the temporal relationship between A and B. Here, the probability that a condition B resulted from a “prime causal suspect” A, out of n known alternative causes plus the influence of coincidence can be represented with the following RR calculation for the causal relationship of \((B|A_i)\):

$$\text{relative risk } (B|A_i) = \left( \frac{\sum_{i=1}^{n} P(B|A_i) - P(B|A')}{\sum_{i=1}^{n} P(B|A_i) + P(B|A')} \right)$$

in which \(P(B|A_i)\) is the probability of the diagnosed condition B given exposure to prime causal suspect \(A_i\); \(\sum_{i=1}^{n} P(B|A_i)\) is the sum of the risk of all known alternative plausible causes of B occurring in the same time frame as \(A_i\), as \(i\) goes from 1 to \(n\) alternative causes; and \(P(B|A_i/t^2)\) is the probability of the coincidental occurrence of B per the time span between prime causal suspect \(A_i\) and the first clinical sign of condition B, designated as \(t\). Taking into account the joint probability of the temporal alignment of (theoretically independent) events \(A_i\) and B, the \(t^2\) factor in the denominator represents the random match probability associated with coincidental cause \(A_i\). The resulting algorithm provides a metric for assessing the relative nature of causal influences while taking into account the biomechanically or epidemiologically derived injury risk values in the numerator of the RR calculation as well as the duration of time between exposure to the prime causal suspect and the injury outcome.

**Example: causal analysis of forensic biomechanics data**

An example of a common application of biomechanical analysis in a forensic setting is the analysis of a claimed intervertebral disk injury following a motor vehicle crash or other trauma. Various features of such injuries can complicate a clinical determination of causation, as both disk and simple spinal strain injuries may initially result in similar clinical presentations. Thus, there is a reasonable appeal to quantifying the biomechanical risk of injury in the collision as an adjunct to clinical causation. The difficulty in such an analysis arises because of the potential for wide variation in the pre-collision condition of the intervertebral disk; variation that results from the nearly universal prevalence of age-related degenerative changes in the disk nucleus and annulus. Additionally, a precise or even rough estimate of the magnitude, direction, and rate of load sustained at a given disk level may be virtually impossible to determine given the number of unknown variables required to accurately reconstruct the intraspinal loads induced by the collision.

If an injury risk can be reliably estimated from a biomechanical analysis, then this probabilistic value could be used in the numerator of the RR calculation once implausibility of the prime causal suspect is rejected; thus \((P(B|A_i) = 0). If there are no known competing discrete traumatic causes of disk injury within the relevant timeframe, then \(\sum_{i=1}^{n} P(B|A_i) = 0\). Thus, the only comparative risk to be included in the denominator is that of the insidious onset of a symptomatic disk derangement that simply coincided with the collision \((P(B|A_i/t^2) = 0)\).

To further develop this example, we hypothesize a rear impact collision with the following facts: the damage to the vehicle was confined to the rear bumper, and the reconstructed speed of the collision was <16 km/h. The driver was a 40-year-old male with a prior history of neck pain and a presumed (non-imaged) cervical \((C5–C6) disk injury three years prior to the collision, but no treatment for the three years prior to the collision. The man develops neck and right arm pain within one hour of the collision, and within one week is diagnosed with a C5–C6 disk herniation based on a magnetic resonance imaging scan. What is the causal RR for the disk injury \((B)\) and the prime causal suspect, the collision \((A_i)\)?

A review of similar collisions and injuries in the National Automotive Sampling System (NASS) of the U.S. National Highway Traffic Safety Administration reveals a very low incidence of cervical disk injury in rear impact collisions of <16 km/h; slightly more than 1 in 2000 collisions or 0.05% [33]. This is not a population-based rate as the NASS database only includes collisions in which at least one vehicle was towed away, and only includes diagnoses that are evident in the first few days following the collision. The actual population-based risk of cervical disk injury under similar crash circumstances is likely 4–10 times greater than what is seen in the NASS database. However, the use of such a low incidence for \(P(B|A_i)\) in the numerator helps avoid a Type I statistical error, in which causation is erroneously found when it is not present. It is important to note that the biomechanical analysis of the crash forces on the occupant is limited to quantification of the collision severity, as there are too many unknown variables to allow for a reliable quantification of the forces at the C5–C6 disk.
For the denominator, the evidence given for the example assumes there are no known competing injury causes outside of the coincidental onset of symptoms within the single hour following the collision, and so \( \sum_{i=1}^{n} P(B_i|A_i) = 0 \). Thus, the probability that the driver would have developed neck pain indicative of a disk injury in the hour following the collision had the collision not occurred is arrived at using the random match calculation described earlier. As an estimate in the role of coincidence, we first must have some idea of the risk of the condition occurring over some period of time \((t)\) during which it is assumed that the man’s risk was stable and constant (rather, there is no evidence that it was not stable and constant). It is reasonable to assume that a 40-year-old man’s risk of developing a symptomatic disk injury would be relatively stable within the period of the year prior to the collision based on his daily driving patterns. Thus the annual prevalence of disk injury in male drivers in this age group can be used for the denominator. Unfortunately, there are no reliable values for the annual prevalence of disk injury; however, the annual prevalence of any episode of neck pain is estimated at an average of 37.2% across a number of studies and populations [34]. Using the prevalence of the far more common condition of neck pain rather than cervical disk injury further reduces the probability of a Type I error. Finally, because the one-hour post collision timeframe is a contribution to the causation assessment, we need to consider any one of the \( t = 8,760 \) hours in a year as a possible coincidental occurrence of the injury. Thus the RR calculation for this biomechanical example is:

\[
\text{relative risk (disk injury | 16 km/h crash) } = \left( \frac{0.05\%}{0 + \frac{37.2\%}{(8760\ h)}} \right) = 103,142
\]

The resulting odds of 103,142 to 1 are the relative risk in favor of the subject collision having caused the disk injury, taking into account the competing explanation of coincidence. For presentation in a forensic setting this is the equivalent of >99.99% probability of causation. From this example it is evident that the temporal proximity between the injury exposure and the relevant symptom onset (here 1 hour) plays a critical role in the assessment of causation because of the random match calculation. For the sake of comparison, if the symptom onset in the example had been delayed by 2 days, the RR odds in favor of causation would have dropped to 45 to 1 or 97.8% as the most probable cause.

The preceding example helps to indicate the limitations in the approach advocated by Walz and Muser and others, when they suggest that knowledge of occupant position, restraint use, etc., provides the primary basis for assessing the probability that a claimed “whiplash” injury actually resulted from a particular collision [14]. Without accounting for the risk of the symptoms indicative of injury absent the collision as well as the temporal proximity between the collision and the onset of symptoms, the biomechanics-only derived injury risk plays a relatively minor role in the determination of cause and effect.

**Conclusion**

Forensic biomechanical analysis is a useful adjunctive tool in forensic medicine, however limitations in its use must be acknowledged and heeded, otherwise the potential for erroneous application arises. When applied to questions of causation, biomechanical assessment of injury risk has a place within a framework of forensic epidemiologic analysis of causation, based upon a systematic approach to causation first laid out by Hill in 1965. These two forensic disciplines (biomechanics and epidemiology) are complimentary and can serve as a reliable means of quantifying the probability of cause and effect in disputed matters.

**Conflict of interest**

Both authors provide forensic consultation in matters that include the application of epidemiologic and biomechanical concepts and analysis to disputed issues.

**References**

A Systematic Approach to Clinical Determinations of Causation in Symptomatic Spinal Disk Injury Following Motor Vehicle Crash Trauma

Michael D. Freeman, PhD, MPH, DC, Christopher J. Centeno, MD, Sean S. Kohles, PhD

Clinical determinations of causation in cases of intervertebral disk (IVD) injury after a motor vehicle crash (MVC) are often disputed in medicolegal settings. No published systematic guidelines exist for making such determinations, which has resulted in infringement by nonclinical personnel into injury causation evaluations, a traditionally clinical activity. The result is causal determinations that are potentially disconnected from clinical observations of injury. The purpose of this review was to evaluate the current literature on causation, causal determinations after trauma and IVD injury after MVC, and to develop a practicable, logical, and literature-based approach to causation determinations of symptomatic IVD injury after MVC. The results of the review indicate IVD injury can result from any MVC regardless of magnitude, thus meeting the first criteria of causation, biologic plausibility. Individual determinations of causation depend entirely on the temporal association between the collision and the symptom onset (the second criterion) and a lack of a more probable explanation for the symptoms (the third). When these causal elements are met, clinicians can assert causation on a “more probable than not” or “reasonable probability” basis. Because of a lack of an established or reliable relationship between collision force and the probability of IVD injury the investigation of collision parameters is not a useful adjunct to causal determinations.

INTRODUCTION

The origin of injury to the intervertebral disks (IVD) of the spine is a common source of dispute in medicolegal circumstances. Whereas treating clinicians typically make determinations of causal relationships between a trauma and an injury primarily based on the patient-related history of the traumatic event and the onset of symptoms, opposing experts frequently rely on population-based assumptions as a basis for disputing causal relationships. This scenario is observed most often when disk injuries and symptoms are attributed to motor vehicle crashes (MVC), because many crash-related injuries result from the negligence of another driver; thus, the nature and extent of an occupant’s injuries may be contested. There is typically a strong third-party (insurer) interest in minimizing compensatory payments on behalf of the individuals who are at fault for the crash. As a result, an adversarial system has developed in the United States and many other industrialized nations, with the injured patient and his or her treating physicians on one side and the party responsible for the crash, his or her insurer, and the insurer’s medical and scientific consultants on the other side [1,2]. Although some clinicians prefer to avoid causal determinations altogether, it can be argued that such determinations are part and parcel of regular medical care. To avoid such a determination solely because it may be later disputed by a party with interests opposing those of the patient is to avoid a duty to the patient; a duty that can have a significant impact on the patient’s well-being. This is not to say that dissemblance and fraud do not exist in the patient population; however, when there is no evidence of dishonesty, there is no reason to assume the presence of fraud simply because a patient is claiming injury after a MVC that was caused by another’s negligence (the most common setting for a legal claim for monetary damages). Such events are, for the most part, randomly distributed in the population that uses motor vehicles and not selective for people...
in whom fraudulent behavior is more common. When allegations of fraud are made in a medicolegal context, despite the absence of evidence of fraud, it is reasonable to question the motivation behind the assertion, because forensic consultants who provide partisan opinions are no more immune to secondary gain pressures than are litigants claiming injury [3].

There are some authors who maintain that for a determination of cause and effect to be made between a diagnosed injury and an MVC, a biomechanical analysis of the forces is necessary to determine whether or not a particular collision had the potential to cause the injuries [4] or that a detailed crash reconstruction is necessary before a causal association can be determined [5]. In contrast, most authors who have documented IVD injuries associated with an MVC make no mention of a biomechanical force analysis or a crash reconstruction [6,7]. The latter approach is most typical of the clinically pragmatic approach to causation. If it is known that an injury can be caused by a trauma, then there is no practical or logical purpose served in quantifying the degree of the trauma in order to assess the risk of injury after the event. Even injuries that are deemed highly unlikely or improbable are by definition still slightly likely or possible, and thus injury presence is typically evaluated after the fact clinically, and not called into question by a post-hoc suggestion of infrequency.

For the purposes of the present inquiry, a Medline search for the years 1980 through 2009 (as of July 2009) using the medical subject headings (MeSH) terms intervertebral disk, intervertebral disk displacement, and disk, herniated as well as the non-MeSH terms cervical disk, in conjunction with “injury” and/or “trauma,” along with “cause,” “causal,” and “causation” in various combinations did not elicit any publications that addressed an organized approach to determinations of causation between disk injury and MVC exposure. It appears that, at the present time, there is no widely accepted scientific or clinical standard for such determinations.

The purpose of this article is to assess the evidence for the causal relationship between disk injury and MVC trauma based on the published literature on causal standards and the science of both disk injury and MVC trauma. Further, the authors propose a systematic and scientifically based approach to individual determinations of causation.

**PRIOR PUBLICATIONS ON CAUSATION**

Determinations of causal association between a noxious exposure (microorganism, chemical, trauma, or other) and a disease or injury outcome in populations is an area of considerable interest to epidemiologists and others who study such issues. Individual clinical observations of cause and effect can serve as a reasonable basis for a case study and suggestion of a relationship. It is, however, difficult to draw generalizable conclusions from such reports. Case series provide stronger evidence of causation; however, without a control group of unexposed subjects, there is no way to determine which effects can be isolated to a particular exposure. Even when case-control studies, in which diseased or injured subjects are compared with healthy subjects for level of exposure to a suspected noxious agent, show positive correlations between exposure and disease or injury outcome, the results cannot be interpreted as validated evidence of a causal relationship, as confounding factors may exist that can obscure true relationships. An example of confounding would be the relationship between coffee drinking and lung cancer. Coffee drinkers may have a higher rate of lung cancer than non-coffee drinkers, but only because smokers are more likely to drink coffee than nonsmokers, not because coffee is causally linked to lung cancer.

In an attempt to more clearly identify causal relationships between noxious exposures and outcomes in populations, Hill outlined a set of 9 criteria that needed to be considered to reasonably conclude that a particular noxious exposure could result in a particular disease [8]. Hill's criteria have served as a kind of seminal gold standard, in that they have been adopted and modified by many others, including the Food and Drug Administration, academic investigators, and pharmaceutical companies, as a means of approaching issues of causation systematically [9]. Based in part on Hill's work, Miller et al proposed a 4-stage process for identifying environmentally caused rheumatic disorders, an application that is more applicable to the goals of this article, because it allowed for assessments of causation in individuals rather than as a population-based construct [10]. McLean subsequently adapted Miller et al's work for causation in instances of fibromyalgia after MVC [11]. All causal criteria can be distilled to a minimum of 3 common and essential elements, which are as follows [12,13].

1. There must be a biologically plausible or possible link between the exposure and the outcome. For example, trauma and fracture are plausibly linked, but trauma and leukemia are not. Plausibility is a low threshold that is exceeded with relatively weak evidence, such as from small observational studies (case studies or case series with small numbers of subjects) or from the results of well-designed experiments with many subjects. Biologic plausibility only pertains to whether an outcome can possibly result from an exposure, and is unrelated to the rate or frequency of the outcome. Thus, evidence of low incidence is not evidence of biologic implausibility.

2. There must be a temporal relationship between the exposure and the outcome. The outcome cannot preexist the exposure; however, the outcome of interest may have preexisted the exposure in a less severe form that was worsened by the exposure. Such a determination requires an accurate documentation of the signs and symptoms of the condition of interest both before and after the exposure of interest. Additionally, the outcome cannot postdate the exposure by a period that is considered, from a clinical perspective, to be too long or too short to relate the two. This determination is highly dependent on the specifics of any case. For example, an injury to a nerve root may cause immediate pain, but electromyogram changes may lag behind by weeks or months. The deter-
1. The patient was an occupant of a motor vehicle at the time of a collision or event.

2. The collision or event transmitted a force through the patient’s spine over a very short time, typically 70-120 ms [19]. Note that the magnitude of the force is not addressed in the definition, because any degree of forceful loading could potentially produce a symptomatic disk injury in a sufficiently susceptible individual.

There are 2 most likely scenarios in which a forcefully loaded IVD may become symptomatic.

1. A healthy disk with little or no changes associated with degeneration is subjected to a significant load resulting in disruption of the annulus and possibly instantaneous migration of nuclear material posteriorly toward the disk periphery and into the epidural space or the lateral recesses. This is the scenario most likely to produce injury to tissue surrounding the disk, including vertebral endplate fracture and ligamentous disruption, and is most likely to be associated with higher energy crashes with significant vehicle damage [20].

2. A degenerated disk (defined as a disk in which one or more of the following are present to some degree: desiccation, fibrosis, narrowing of the disk space, diffuse bulging of the anulus beyond the disk space, anular tears, endplate sclerosis, and osteophytes at the vertebral apophyses [14]) is loaded in a manner that initiates a symptomatic response. A significant proportion of the asymptomatic population has some degree of disk degeneration as seen on MRI; Matsumoto et al have performed the largest study to date on the topic, with MRI scans of the cervical spines of 497 asymptomatic subjects [21]. These authors describe a progressive direct relationship between the presence and severity of degenerative changes and age, with the disk degeneration noted in a range of 17% men and 12% of women in their 20s to 86% and 89% of men and women older than 60 years. Other authors have described similar findings with broad ranges of values depending on the age of the subjects; from virtually no disk degeneration in subjects <30 years of age, to 14% in those 30-40, and 62% in subjects >40 years of age [22,23]. Similar observations have been published with regard to lumbar spine degeneration among asymptomatic patients. Jensen et al described a cohort of 98 subjects who underwent lumbar MRI, noting that only 36% of subjects had no abnormal disks. Fifty-two percent of the scans were interpreted as having a bulge at a minimum of one level, 38% had abnormality at more than one level, 27% had a protrusion, and 1% had an extrusion [24]. These findings are consistent with those reported by other authors [25].

**A Causation Algorithm for Disk Injury after MVC**

In this section, the authors present a practical guideline for clinical determinations of causation in symptomatic disk injuries after an MVC, based on the application of the aforementioned 3 causal elements to the facts and findings in an individual case. The following definitions and assumptions are made. An IVD injury is defined by all of the following characteristics [14,15] (for this definition, “injury” is used only to define a disk that has become symptomatic after a discrete loading event of any magnitude, and the cause of the injury is not addressed).

1. It is symptomatic. This may include classic symptoms of segmentally appropriate radiculopathy, regional axial pain, or diskogenic referral of pain in a nondermatomal distribution.

2. There is clinical imaging evidence (magnetic resonance imaging [MRI], computed tomography, or diskography) that the symptomatic disk fits the generally accepted definition of a degenerative-traumatic disk [14]. This includes anular tears, herniations, and degeneration. Although there are findings specific to disk trauma, such as vertebral endplate fracture and edema presence in the marrow adjacent to the endplate, many traumatically initiated disk injuries are indistinguishable from degenerative changes in the disk [16]. Because of mediocre sensitivity and specificity (ie, true- and false-positive rates), imaging findings alone are insufficient evidence for either a causal or noncausal determination in a case of suspected or disputed traumatic disk injury [17,18].

MVC trauma is defined as follows.

1. The patient was an occupant of a motor vehicle at the time of a collision or event.

2. Biologic Plausibility. Can an MVC of any severity cause a disk injury? Crash testing of intact cadavers at accelerations recorded for no-damage rear-impact collisions has demonstrated IVD injury that cannot be detected by conventional
imaging such as CT and MRI [26]. Although the biomechanical literature on experimental loading of disks in cadaveric sled testing has demonstrated disk injuries even in relatively low accelerations (3.3 and 4.5 times the force of gravity or ‘g’ [26]), such ex vivo testing of healthy disks does little to demonstrate minimal thresholds for injury in live human populations exposed to real-world crashes. No cadaveric testing can duplicate the significant variation in the manner in which loads are transmitted to the human spine in a MVC, given the variation in occupant position, restraint systems, vehicle interiors, and muscle loads. Additionally, the potential for variation in the condition of an individual IVD, including the ability of the disk to resist forceful loading from an external source, is too large to ever define in an experimental setting.

A relevant literature search revealed a complete absence of any publications suggesting that symptomatic IVD derangement cannot result from MVC exposure of any magnitude. To the contrary, injury to the disks of the cervical and lumbar spine has been documented in the relatively low level accelerations associated with little to no-damage collisions, roller coaster rides, and even sneezing [27-30]. It is reasonable to conclude, as a general precept, that the forceful loading of the spine that can occur in any MVC is a biologically plausible (possible) cause of symptomatic disk injury.

Temporal Association. Temporality is the strongest evidence of causation in evaluating the patient with post-MVC disk injury. It must be first established that the MVC preceded the onset of symptoms attributed to the disk injury. The exception to this rule is when a previously symptomatic disk is exacerbated by an MVC to the point that the course of care is significantly altered (ie, a previously diagnosed nonsurgical disk condition becomes surgical directly after an MVC). In such cases, the determination must be made clinically as to whether the disk symptoms were likely to have worsened to the point that the surgical intervention was inevitable absent the MVC. The mere fact that the disk was previously symptomatic is not sufficient to draw the conclusion that it would have required surgery, given the good outcomes for patients with diskogenic symptoms who use a variety of surgical and nonsurgical treatments [31-33]. The symptom onset must be in reasonable proximity to the time of the MVC; however, disk injury symptoms may initially present as identical to symptoms of spinal strain or sprain, and it may be weeks or months before an MRI is performed and the disk injury diagnosis is first seen [34]. Additionally, disk injury symptoms may be progressive, as an injury to the annulus may progressively allow nuclear migration toward neurologic structures over a period of time while the patient is in weight bearing and active. Determinations as to whether delayed symptoms fulfill the temporal association criteria must be made by clinicians on a case-by-case basis.

An obvious weakness of the temporality criterion is that it can only be established from the history given by the patient, and patient histories can be erroneous or falsified. The former is unlikely in the case of injury, however, because one of the hallmarks of injury is that there is a close temporal relationship between cause and effect (this lack of clear temporal proximity in repetitive traumatic exposures is the reason that injuries such as carpal tunnel syndrome are classified as diseases) [35]. This close temporal relationship means that it is unlikely that a patient will misattribute symptoms to a traumatic cause. As a practical matter, the most reliable indicator of when a patient began to have symptoms after a trauma is the patient. This is not to say that patients do not get details concerning an injury event wrong; patient perception or recollection of vehicle speeds, crash sequence, and other details of a collision may be inconsistent with the facts of a case for reasons other than untruthfulness. For example, an occupant who is injured in a minimal damage rear end collision may believe that the impact must have been at high speed despite a lack of physical evidence that this was the case.

With regard to the truthfulness of the patient, it is not the role of the clinician to investigate the veracity of the history given by the patient. Absent some ancillary indication of deception, clinicians are generally safe to believe their patients. The rate of fraud of any kind in auto insurance claims for treatment benefits (personal injury protection coverage) was estimated to be 6% for 2007 [36]. Even if every instance of fraud in auto insurance cases consisted of a patient giving a false history, clinicians would still be justified in believing their patients 94% of the time.

Intuitively, it makes sense that the closer the onset of symptoms is to the time of the collision the stronger the causal relationship between the MVC and the disk injury, and the strength of this relationship can be quantified using an indirect approach. For example, in a case of a 40-year-old male with neck pain and cervical radiculopathy that arose within 12 hours of a rear impact collision (later attributed to MRI findings of disk herniation), assuming no intervening trauma, it can be postulated, as a tautology, that either the crash caused the disk injury or that it was coincidental to the collision, regardless of cause. Thus, the probability that the crash caused the disk injuries \( P(\text{disk}_{\text{MVC}}) \) plus the probability the disk symptoms are coincidental to the crash \( P(\text{disk}_{\text{COINC}}) \) when added together account for all of the possible causes (100% or 1) of the disk injury and associated symptoms. If \( P(\text{disk}_{\text{MVC}}) + P(\text{disk}_{\text{COINC}}) = 1 \) then rearranging the terms gives \( 1 - P(\text{disk}_{\text{COINC}}) = P(\text{disk}_{\text{MVC}}) \), meaning that if the probability that a disk injury occurred coincidentally on the same day of the MVC can be determined then the probability the MVC caused the disk injury can also be determined indirectly.

The probability of 2 unrelated events occurring in close temporal proximity can be calculated by multiplying the probability or odds of one times the other (odds are the ratio of 2 probabilities). For example, the odds of rolling 2 sixes in a row with a fair die are \((1:6) \times (1:6) = 1:36\). This is because there are 36 possible combinations for any 2 rolls, including 1 and 1, 1 and 2, 1 and 3, etc, and only one of the combinations is a 6 and 6.
In a similar fashion the odds of a disk injury occurring coincidentally on the day of a collision can be calculated for the example given earlier. The annualized odds of involvement in a MVC for a 40-year-old male are 1:14 (once every 14 years), and the daily odds of crash involvement are approximately 1:5100 (one crash every 5100 days) [37]. The odds of the spontaneous development of disk symptoms are a bit more difficult to estimate. For example, it can be said that the patient had a single-day episode of neck and arm pain 12 months before the MVC, with no intervening episodes until the time of the collision. Thus, the odds that the symptoms would recur on any single day were no greater than 1:365. Multiplied together these 2 odds yield an odds of a disk injury occurring coincidentally on the day of the collision of 1:1,861,500. Conversely, the odds in favor of a causal relationship between the collision and the disk injury, given the onset of symptoms on the day of the collision, are 1,861,500:1.

**Lack of Likely Alternative Explanations**

Alternative explanations for a disk injury include an intervening trauma that followed the MVC and preceded the symptoms indicative of disk injury. Another alternative explanation is the insidious onset of symptoms, possibly associated with a trivial perturbation such as sneezing. Although minor forces are possible causes of disk injury, they cannot be preferentially selected over the substantially greater forces of a collision when the two have occurred in close temporal proximity. If, for example, a patient sustains what is diagnosed as a lumbar strain in a collision and a week later sneezes and has a sudden onset of radicular pain associated with a disk extrusion, then it is most likely that the disk was injured in the MVC and the sneeze only prompted a progression of symptoms. It is not reasonable to point to minor forces associated with daily activities as a likely cause of a symptomatic disk in preference to the significantly greater trauma associated even with minimal and no-damage MVCs that is temporally relevant to the symptoms, particularly when the patient performed such minor activities with no difficulty prior to the MVC. A “likely” alternative explanation is one that is most probably causal and the best explanation to fit all of the facts at hand, not one that is only a possible cause with no historical or temporal evidence to link it to the symptomatic disk injury.

**DISCUSSION**

Individual determinations of causation for LID injury after MVC by clinicians require only that temporal association and lack of likely alternative explanation be satisfactorily present. The biologic plausibility or possibility that symptomatic disk derangement can follow any degree of forceful loading of the spine is satisfactorily present and need not be revisited in individual cases. Thus, the concept of measuring crash forces in detail as a means of determining disk injury potential can be abandoned as serving no useful purpose in causation determinations.

There is no reliable index of the degree of force required to cause a particular degree of disk derangement. Walz and Muser theorized that unless a crash reconstruction and biomechanical loading assessment was performed for a particular crash and occupant that causation could not be determined, or, put another way, it must be determined that a given crash was sufficiently forceful to cause diagnosed injuries [4]. These authors suggested that parameters such as seat belt use, head restraint and seat properties, age, body size, and preexisting damage (sic) to the spine, inter alia, be taken into account when making causation determinations, effectively taking causal determinations out of the hands of clinicians. What these authors do not explain is how an evaluation of any or all of such factors could help determine whether or not any type or severity of injury is possible after a crash. Even if it could be determined that the risk of disk injury was exceedingly low for a particular occupant in a particular crash, this would not be evidence that contradicts any of the 3 elements of causation. Thus, this biomechanical risk model of injury presence has no utility for real-world determinations of causation.

After a determination of causation has been reached, the legal standard for expressing the opinion is as “more probable or likely than not” or as a “reasonable probability” or “reasonable medical probability” [38]. In some jurisdictions, the standard is that the clinician must be “more than 50% certain” that the opinion is correct. The purpose of such language is to describe the results of an internal process of weighing evidence, and to arrive at a conclusion that the clinician is more certain than not that his or her opinion is accurate or true. The methodology presented in this article is designed to provide a framework for clinicians to arrive at opinions of causation that can be expressed in terms of what is more likely than not. Thus, when using these guidelines in a narrative report setting, the clinician can write that the essential causal elements of biologic plausibility, temporality, and lack of likely alternative explanation have been met for a given case (if true), and that a particular symptomatic disk injury resulted from a particular motor vehicle crash, as a reasonable medical probability.

It must be noted that the model of causation presented herein suffers from the fact that there is no other alternative against which it can be evaluated for accuracy. Clinicians receive little or no formal education in medical school or residency regarding a systematic approach to causation. Thus, the authors recommend that guidelines presented herein be evaluated, modified if necessary, and considered for adoption by consensus by appropriate scientific and medical organizations.

**CONCLUSIONS**

Individual determinations of causation for disk injury after MVC are most appropriately conducted by clinicians, based primarily on an evaluation of the temporal association be-
between the MVC and the symptom onset. The lack of a valid or meaningful collision force threshold below which it can be said that a disk injury will not occur means that investigation of collision parameters for the purposes of injury causation assessment is a pointless endeavor.

REFERENCES