

# Modelling of links from a laboratory test result to real-world performance: The case of pedestrian collisions

T. P. Hutchinson, R. W. G. Anderson, and D. J. Searson

*Centre for Automotive Safety Research, University of Adelaide, South Australia 5005  
Email: paul@casr.adelaide.edu.au*

**Abstract:** *Background.* Hardware, software, and people are often tested in one set of conditions, but are expected to perform under many different circumstances. For example, consider pedestrian headform testing. An approximate sphere, with an accelerometer inside it, is projected at the front of a car. The speed is specified, but real pedestrian impacts are at a wide range of speeds. The headform mass is specified, but real pedestrians have a range of effective head masses. The acceleration trace is summarised in order to check that the car is not overly injurious if a pedestrian is struck. This paper will set out the principles of calculating the real-world consequences --- that is, averaged over the range of speeds and effective head masses --- of a particular test result under specified test conditions. Such a calculation is not common in the road safety world, and we do not think it is common in other testing contexts. It is a specific example of the more general problem of using performance in particular test conditions to estimate average real-world performance. (But note that the discussion is not primarily about going from one condition to another, such as from harsh conditions in which failure is accelerated to normal conditions in which failure is infrequent. At least in our context, the test conditions are realistic. The issue is rather that of averaging over a variety of conditions.) *Proposed procedure.* It is proposed to calculate an estimate of the average level of performance --- averaged over different conditions, that is. This calculation has three components. (1) An equation for the dependence of performance on conditions. (2) An equation for the cost of (i.e., how bad or good are) different levels of performance. (3) The probabilities of different conditions. The three components come together in a summation or integration that represents the averaging over different impact conditions. *Applications.* The equation permits, for example, the calculation of the changes that result if test performance is improved, or the probabilities of different conditions change. This paper will present the specific methods we have developed, and will suggest they can be expressed in quite general language. Our interest is in impact testing, but the core issue --- the implications for average real-world performance of a test in one set of conditions --- must be in the minds of people concerned with a great variety of tests.

**Keywords:** *Average severity, Cushioning, Impact testing, Pedestrian subsystem testing, Speed of impact*

## 1. INTRODUCTION

A tremendous amount of testing is conducted: testing of things, software, people, and teams of people. A system of testing presumably exists because of its real-world relevance, yet there is a contrast between the test being conducted in tightly-specified conditions and the real world being highly variable. This paper comes within the area of testing (of things or people) viewed in its generic aspects, or as a system. It will set out the principles of calculating the real-world consequences of a particular test result obtained in specified test conditions.

The example that provides the starting point is from impact testing, and specifically pedestrian headform testing. From a pedestrian's viewpoint, the exterior of a vehicle (in particular, the bonnet) is a cushion to protect the pedestrian against the much stiffer structures underneath (Hutchinson et al., 2011). In the testing, an approximate sphere, with an accelerometer inside it, is projected at the front of a car. Individual tests are at particular locations on a car's front, and the basic unit of analysis is a location, though a summary score for the whole car is what is publicised. The speed is specified, but real pedestrian impacts are at a wide range of speeds. The headform mass is specified, but real pedestrians have a range of effective head masses. The present paper aims to demonstrate the broad applicability of the core ideas of the procedure developed in that specific context.

Section 2 will develop the method. Details are given by Hutchinson et al. (2012a), and other authors have also advocated generalising to a range of conditions (Horsch, 1987; Korner, 1989; Searle et al., 1978; Viano, 1988). Section 3 will examine applications (starting with impact testing and then extending into other fields), and section 4 will discuss what has been presented.

## 2. EQUATION FOR AVERAGE COST

Three inputs will be needed to the calculation.

- The probabilities of whatever it is that varies in the real world.
- The utility or cost corresponding to any result that might be observed in a test. Due to lack of information, the utility or cost may have to be approximated by the probability of something (such as a level of injury) occurring.
- The quantity that is measured in the test.

Returning to the specifics of pedestrian head impacts, the probabilities refer to impact speeds; cost might be represented by the probability of death; and the test output is a quantity termed HIC (the Head Injury Criterion), that summarises the acceleration trace over the milliseconds of impact.

The following notation will be used.

- Let the important single dimensional quantity that varies in the real world be  $x$ , and its probability density function (p.d.f.) be  $f(x)$ .
- Let the cost or utility be  $p$ . The symbol  $p$  is used because this may be the probability of something.
- Let the quantity that is measured in the test be  $h$ . This depends on  $x$ , and it determines  $p$  (it is assumed that  $h$  includes all the information needed to determine  $p$ ). Thus the notation needs to be  $h(x)$  and  $p(h)$  or  $p(h(x))$ .
- Furthermore, the purpose of testing is to measure the thing being tested. Let this be labelled as  $i$ . More fully, then, the notation is  $h(x, i)$  and  $p(h(x, i))$ .
- The test is carried out at a specified value of  $x$ ,  $x_1$ , and the observed result is  $h_1$ .

Theory is needed to infer the function  $h(x)$  from single observation of  $(x_1, h_1)$ . For example, theory may require that  $h(x)$  is a straight line of a specified slope, and knowing  $x_1$  and  $h_1$  determines exactly what that straight line is.

The frequency of events at  $x$  is the total number multiplied by the p.d.f. The cost of each is  $p(h)$ . Therefore the cost of all events at  $x$  is proportional to the product  $f(x).p(h)$  --- or, more fully,  $f(x).p(h(x, i))$ . The total cost of events at all the various values of  $x$  will be obtained by summing over  $x$ . The result might be called  $Av(p_i)$ , meaning the average of  $p$ , when what is tested is labelled  $i$ .

In standard mathematical notation, this averaging is written as an integral:

$$Av(p_i) = \int f(x).p(h(x, i)).dx \tag{1}$$

As already mentioned,  $h(x)$  is probably obtained by theory-based generalisation from the test result  $(x_1, h_1)$ . The p.d.f.  $f(x)$  is a description of the relative frequencies occurring in the real world, and  $p(h)$  is a description of how cost depends on  $h$ .

Figure 1 summarises the strategy of this set of calculations in the case of pedestrian head impact testing. For the test speed, the corresponding HIC is needed. From  $(x_1, h_1)$ , theory is used to infer the whole function  $h(x)$ . Then  $h$  is converted into cost  $p$ . Finally, the relative frequencies of different speeds are used to calculate average cost.

Much of the strategy is common sense, and variations are easily imagined. (a) There may be several quantities that vary in the real world, not only one (see Hutchinson et al., 2013). (b) The function  $h(x)$  might be estimated not by theory but instead by doing several tests at different  $x$ . Of course, the disadvantage is the cost of testing. (c) Or  $h(x)$  might be estimated not by a simple formula but instead by some complicated simulation.

Something that performs better than another in the test may not perform better in other circumstances. That is, theory may suggest that different examples of  $h(x)$ , say  $h(x, i)$  and  $h(x, j)$ , will cross over as  $x$  increases. In the case of pedestrian impact, a car bonnet is relatively soft, but “bottoming out” refers to this deforming so much during impact that something very stiff underneath is contacted, and the pedestrian consequently sustains very high accelerations. A stiffer bonnet is worse for low speed impacts, but may be better for high speed impacts because bottoming out is prevented or mitigated. In the field of human performance, one person may score better than another in normal conditions, but the second person may be less impaired by stressful conditions.

Figure 2 illustrates that if  $f(x)$  is represented by distribution A (with most of the distribution being in the range of  $x$  for which the performance of  $i$  is higher), average performance is likely to be higher for  $i$ , but if  $f(x)$  is represented by distribution B, average performance is likely to be higher for  $j$ . Two major complications not illustrated in Figure 2 may occur. (a) The lines  $i-i$  and  $j-j$  are unlikely to be straight. Furthermore, they may be estimated on the basis of one datum point and a theory that is of questionable validity. (b) The lines  $i-i$  and  $j-j$  may not show a cost or utility that can be averaged in a meaningful way. The measure of performance,  $h(x)$ , needs to be transformed into  $p$ .

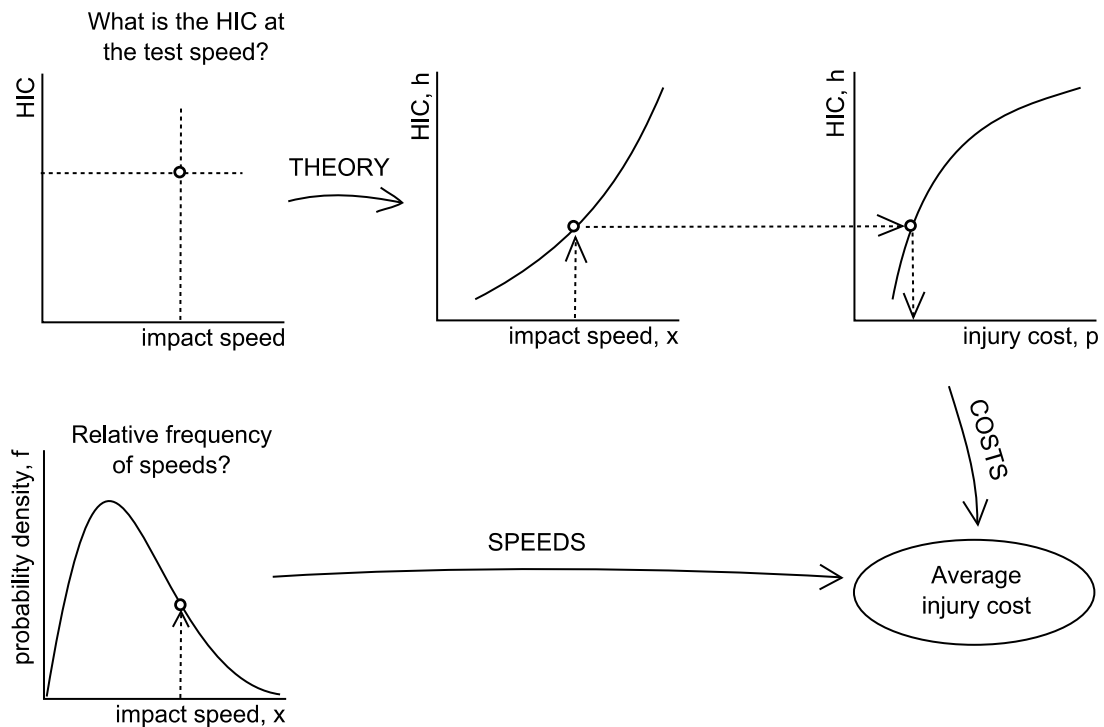
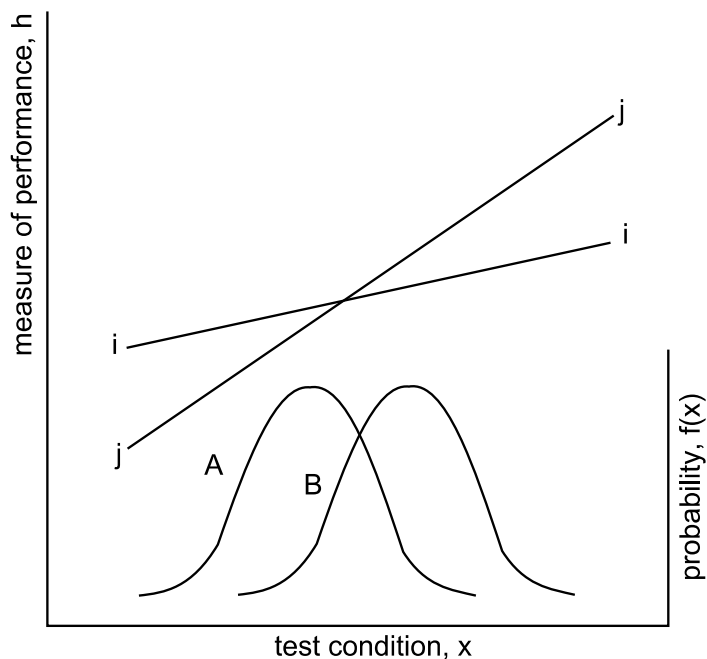


Figure 1. Strategy of calculating average cost using Equation (1).



**Figure 2.** Cross-over of the functions  $h(x)$  that apply to two items,  $i$  and  $j$ .  $A$  and  $B$  are two alternative distributions  $f(x)$ .

### 3. AREAS OF APPLICATION

#### 3.1. Testing of Things

As well as pedestrian headform testing, a number of other methods of impact testing are used in safety contexts, including a helmeted headform hitting a rigid anvil, a projected impactor hitting a helmeted headform, and a dummy representing the whole human. The methods developed for a headform projected against a car’s exterior will be relevant in these other circumstances, but may need some modification (e.g., in respect of the theory for inferring  $h(x)$  from  $(x_1, h_1)$ ).

Testing is also conducted of protection for the limbs and body. Helmets and body armour may be sufficiently effective that impact from a projectile is converted to blunt impact of sufficient severity to be dangerous itself. And in some contexts, the concern is not severe sudden injury but instead repeated impacts causing cumulative injury (e.g., for footwear and keyboards).

There are many other contexts in which the mitigation of impact is important, and where there is a contrast between testing being conducted at a standardised level of violence whereas real-world impacts occur at many different levels of violence. The principles and methods proposed here are potentially transferable to contexts such as the following (Hutchinson et al., 2012b): handling and packaging of manufactured items, whether large and relatively robust or small and delicate; handling and packaging of fruits and vegetables; manipulation of items by robots (if one component is to be fitted to another, it is usually desirable that initial contact is not violent); inadvertent impact of robots with humans; impacts of birds with aircraft; undesirable powdering of bulk materials during transportation and processing.

Equation (1) may also be relevant to other forms of testing. An imaging system may be tested at one level of illumination and for viewing one distance, yet be used in a range of illumination conditions and for viewing a range of distances. The same applies to listening systems, and to many other types of hardware. Software interacts with other software, with hardware, and with people. The characteristics of any of these may be much more variable in use than under test conditions.

### 3.2. Testing of Individuals and Groups

If the slope of  $h(x, i)$  is approximately the same for all  $i$ , calculation of  $Av(p_i)$  and  $Av(p_j)$  has the benefit of quantifying the comparison between  $i$  and  $j$  --- this may be useful in itself or as input into a decision along with other considerations such as (for example) the cost of  $i$  and  $j$ . Alternatively, if the slope of  $h(x, i)$  does depend on what  $i$  is, it will be vital to make use of Equation (1), or otherwise outcomes at different  $x$  cannot be weighted appropriately. Crossover of two functions  $h(x, i)$  and  $h(x, j)$  can arise easily, as is shown by the examples below.

The function  $h(x)$  is a generic measure of performance. In impact testing,  $h$  is usually a measure of how severe the impact is: the smaller, the better. In other contexts, it is usual for performance to be measured in such a way that the higher, the better, and that language will be used in this section.

*Floor and ceiling levels of performance.* A “floor” and a “ceiling” to  $h(x)$  refer to levels of performance in, respectively, the most difficult and the easiest conditions. It is common to think that someone of higher ability than another, as evidenced by better performance on tasks of average difficulty, will be better at difficult tasks of the same type and at easy tasks of the same type. But this is an incomplete account if people may differ in respect of their floor level of performance and their ceiling level of performance, as well as in their typical level of performance.

- Consider easy tasks that almost anyone can succeed in. Whether or not these are actually successfully performed may depend largely on characteristics usually thought to be part of personality rather than ability: being attentive and careful and avoiding silly mistakes.
- Consider difficult tasks that almost everyone will fail at. The degree to which these are successfully performed may again depend largely on personality characteristics: willingness to have a go, creativity, psychological acceptance of partial success, perseverance, persistence, and so on. (In many testing situations, success may also occur from random guessing.)

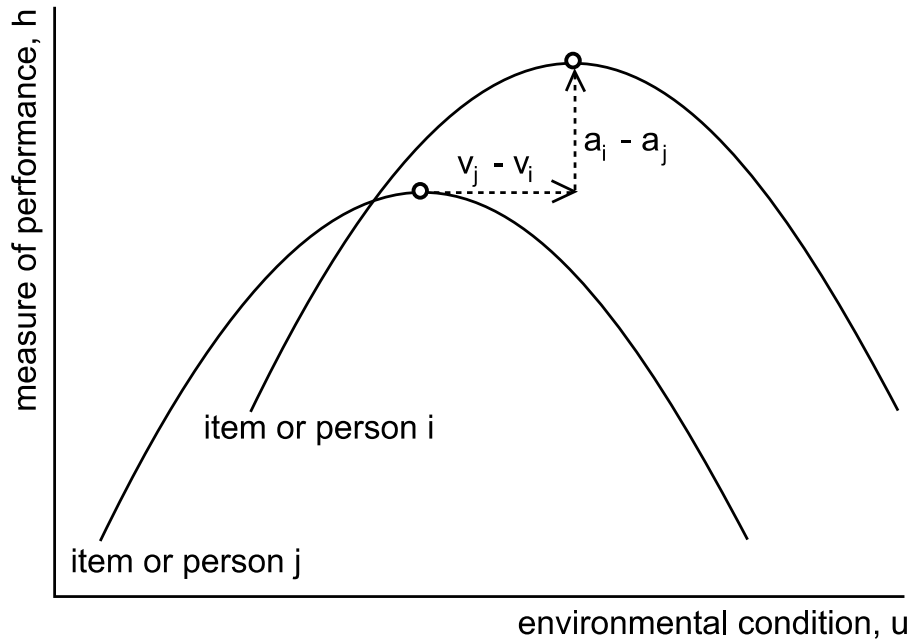
Thus  $h(x, i)$  (referring to person  $i$ ) and  $h(x, j)$  (person  $j$ ) may cross over. The relative proportions of real-world tasks that are of average difficulty, are easy, or are difficult will determine how important are the three characteristics. Names for the three characteristics might be ability, carefulness, and persistence. In such a case,  $h(x, i)$  would need to be obtained from three test results  $(x_1, h_1)$ ,  $(x_2, h_2)$ , and  $(x_3, h_3)$ , or perhaps from  $(x_1, h_1)$  together with measurements by quite different methods of carefulness and persistence.

*Probabilities and proportions.* If performance is expressed as a probability or proportion (of succeeding in a task, for example), it will necessarily have both a floor and a ceiling, as probabilities are between 0 and 1.

*Optimal arousal.* Stress and similar concepts are important examples of conditions. Human performance is commonly said to be best when stress or arousal or stimulation is moderate, and lower both when one is too sleepy or too excited. That is, stress / arousal / stimulation has an inverted-U effect on performance. Testing people at one value of arousal will not give sufficient information about how they perform across the range of levels of arousal, or what their average performance will be.

*Matching of the task to the person.* More broadly than optimal arousal, performance may be a decreasing function of how different the task is from the task the person is best at. Expressing this generally, suppose that (a) the shape of  $h(x)$  is nonlinear, (b)  $i$  has an effect that is additive with  $h(x)$ , and (c) there is a further effect of  $i$  because what determines  $h$  is actually the sum of an environmental variable  $u$  and a quantity  $v_i$  that is different for different people. In other words, there is some nonlinear function  $g(\cdot)$ , along with quantities  $a_i$  and  $v_i$  that are characteristic of  $i$ , such that  $h$  is  $a_i + g(u+v_i)$ . Then the slope of  $h$  at any given  $u$  will depend on  $i$ . This is illustrated in Figure 3. When environment has different effects on different people, there is said to be interaction between environment and person. Interaction arising from a nonlinear function of a sum has been discussed by Hutchinson (2009).

*Heat strain of humans.* Above,  $x$  was thought of as task difficulty or arousal. It might instead be some aspect of the environment. Temperature, for example, is important to humans and to human performance. People and their clothing (e.g., protective clothing) may be tested at one temperature, but it is expected that they do not suffer hyperthermia and do perform their task satisfactorily over a range of temperatures. Heat strain models have been developed to predict body temperature from such factors as metabolism (which in turn will depend on body weight, weight of load, intensity of exercise, and so on), the environment (dependent on temperature, the insulation of the clothing, and so on), and available cooling (dependent on the permeability to water vapour of the clothing, humidity, wind speed, and so on). Such models could be viewed as analogous to  $h(x)$ .



**Figure 3.** Cross-over as a consequence of a nonlinear function being moved horizontally as well as vertically.

#### 4. DISCUSSION

##### 4.1. Looking Ahead: Sensitivity Analysis

Testing is often conducted within a setting of decision-making. For example, the issue may be whether performance is good enough or not, or which of several items tested is the best. The setting being decision-making, some forms of gross error in specifying the required functions may be tolerable, as the error may apply to all items approximately equally. This is fortunate, as the task of specifying the functions may be very difficult. The question will arise of how big an error will make how much difference to the end result.

We suggest it may be feasible to write software for sensitivity analysis of Equation 1 that could be used in any context, hardware or software or human or organisational. (a) The heart of such software would be the computation of the result using Equation 1, and naturally there would need to be specification of the required functions  $h(\cdot)$ ,  $p(\cdot)$ , and  $f(\cdot)$  as inputs, so that the change in the result from a change in these functions could be determined. (b) Some effects of the specification of  $h(\cdot)$ ,  $p(\cdot)$ , and  $f(\cdot)$  cannot be known unless the population of items under test is specified also. For example, it is common sense that if two items  $i$  and  $j$  are very similar in the sense that the functions  $h(x, i)$  and  $h(x, j)$  are close together, a decision on which is the better will be sensitive to what  $p(\cdot)$  and  $f(\cdot)$  are. Consequently, the computation of Equation 1 and the specification of the required functions will need to be set within a specification of what the test items are. (c) Finally, there will need to be specification of the form of the results. Examples would be the actual absolute values calculated by Equation 1, comparisons of these with a predefined value representing the threshold between pass and fail, comparison of one item with another, and the rank ordering of several items.

##### 4.2. Concluding Comments

Equation (1) is demanding of data, and may be perceived in many circumstances as too difficult to use. Nevertheless, this procedure is basically a common sense one: performance is worth knowing in a variety of conditions (whether it is estimated by extra testing or theoretically), the direct measure of performance ( $h$ ) needs to be converted into a cost or a value ( $p$ ), and the frequencies of different conditions  $f(x)$  must be considered alongside the levels of performance  $h(x)$ . Improved knowledge of  $h$ ,  $p$ , or  $f$  will all give better insight into what happens in the various real-world conditions, even if use of Equation (1) is only approximate.

Systems of testing have a number of aspects in common.

- The overall aim, of making a transition from testing in a set of prescribed conditions to the real world where conditions vary. An average seems desirable. Ideally, this would be correct in an absolute sense. That may be too ambitious an aim, but an average that is correct in a comparative sense (that is, comparison with a predefined standard, or with another item tested) may be sufficient for purpose.
- The need to generalise from performance in test conditions to other conditions.
- The need to convert performance to something that can be meaningfully averaged (value or utility or cost).
- The need to know the probabilities in the real world of various conditions.

It is concluded that Equation (1) may be applicable to very many forms of testing --- to testing of humans, organisations, and software as well as to hardware.

#### ACKNOWLEDGEMENTS

The Centre for Automotive Safety Research, University of Adelaide, is supported by both the South Australian Department of Planning, Transport and Infrastructure and the South Australian Motor Accident Commission. The views expressed in this paper are those of the authors, and do not necessarily represent those of the University of Adelaide or the sponsoring organisations.

#### REFERENCES

- Horsch, J.D. (1987). Evaluation of occupant protection from responses measured in laboratory tests. SAE Technical Paper 870222, Society of Automotive Engineers, Warrendale, PA.
- Hutchinson, T.P. (2009). Interpretation of data showing something has one effect sometimes and a different effect in other circumstances: Theories of interaction of factors. *ASOR Bulletin* (Australian Society for Operations Research), 28(3), 25-29.
- Hutchinson, T.P., R.W.G. Anderson, and D.J. Searson (2012a). Pedestrian headform testing: Inferring performance at impact speeds and for headform masses not tested, and estimating average performance in a range of real-world conditions. *Traffic Injury Prevention*, 13(4), 402-411.
- Hutchinson, T.P., R.W.G. Anderson, and D.J. Searson (2012b). Testing in order to measure the protection against impact of people, manufactured items, and agricultural produce: How to consider all severities of shock. *In Proceedings of the 7th Australasian Congress on Applied Mechanics*, edited by A. Kolousov, R. Das and S. Wildy pp. 925-932. National Committee on Applied Mechanics, Engineers Australia, Barton, ACT, Australia.
- Hutchinson, T.P., R.W.G. Anderson, and D.J. Searson (2013). An equation for generalising from impact test performance to real-world crashes. Paper 13-0355, 23rd Enhanced Safety of Vehicles Conference. <http://www-nrd.nhtsa.dot.gov/pdf/esv/esv23/23ESV-000126.PDF>
- Hutchinson, T.P., D.J. Searson, R.W.G. Anderson, J.K. Dutschke, G. Ponte, and A.L. van den Berg (2011). Protection of the unhelmeted head against blunt impact: The pedestrian and the car bonnet. *In Proceedings of the Australasian Road Safety Research, Policing and Education Conference*, <http://acrs.org.au/publications/conference-papers/>
- Korner, J. (1989). A method for evaluating occupant protection by correlating accident data with laboratory test data. SAE Technical Paper 890747, Society of Automotive Engineers, Warrendale, PA.
- Searle, J.A., J. Bethell, and G. Baggaley (1978). The variation of human tolerance to impact and its effect on the design and testing of automotive impact performance. SAE Technical Paper 780885, Society of Automotive Engineers, Warrendale, PA.
- Viano, D.C. (1988). Limits and challenges of crash protection. *Accident Analysis and Prevention*, 20(6), 421-429.