

# **Safety in Geometric Design Standards.**

Ezra Hauer  
Professor (Emeritus), Department of Civil Engineering  
University of Toronto

Toronto, December 15, 1999

# Safety in Geometric Design Standards.

Ezra Hauer, 35 Merton Street, Apt, 1706, Toronto, ON., M5S 3G4.  
Hauer@civ.utoronto.ca.

Designers of roads believe that roads built to standards are safe. Lawyers and judges assume that roads designed to standards are appropriately safe. Beliefs, no matter how passionately held, and assumptions, no matter how repeatedly applied, are fallible guides to truth. The truth is that roads designed to standards are *not safe, not unsafe, nor are they appropriately safe*; roads designed to standards have an unpremeditated level of safety. This is the claim to be substantiated

In the first part of this paper I will appeal to common sense and logic. However, to use logic against strong beliefs is like firing lead shot to sink a battleship. This battleship is not built of steel but of trust; a trust that the succession of standards committees that formulated and improved design standards, did so on the basis of factual knowledge about how their decisions affect safety. The second part of this paper aims to diminish this trust. Here I will recount the evolution of three important design procedures in the geometric design standards: those governing the design of vertical curves, those that pertain to the choice of lane width, and those that apply to the design of horizontal curves. The lessons drawn from these historical anecdotes direct attention to the design paradigm that has been shaped by the history and culture of civil engineering. I will argue that the prevailing design paradigm of civil engineering is deficient when it comes to designing safety into roads. The last part of the paper examines options for reforming road design; what to do to so that the roads we build are appropriately safe.

## **1. Roads designed to standards are not safe, unsafe, or appropriately safe.**

When the safety of a newly built freeway was publicly questioned by police officers, the Ontario Minister of Transport proclaimed that the new freeway meets the current design standards and is therefore safe. I think that many road designers would concur. What did the Minister actually mean by 'safe'?

To avoid sterile disagreement, one has to be clear about what 'Road Safety' means. Of two alternative highway designs connecting points A and B and serving the same traffic, that highway design which is likely to have fewer and less severe crashes is the safer one. Thus, the safety of a road is measured by the frequency and severity of crashes expected to occur on it. If so, safety of a road is always a matter of degree. A road can be safer or less safe.

The Minister of Transport could not have meant that the new freeway will be safe because it will be crash free forever. No road in use is crash free. Therefore the 'crash-free' interpretation of 'safe' is of no use and need not be pursued. Perhaps the Minister meant that a road designed to standards is *as safe as it can be*. The usual corollary of such a belief is that when crashes occur on

a road that is as safe as it can be, the drivers are to blame for their misfortune. Perhaps the Minister meant that, even if not as safe as it can be, the new freeway *is as safe as it should be*. It would then follow that the roads that meet standards are 'appropriately safe'.

This historical anecdote serves to focus on two possible relationships between design standards and safety:

- a. Roads designed to meet current standards are as safe as they *can* be;
- or,
- b. Roads designed to meet current standards are as safe as they *should* be.

The verity of statement 'a' is suspect at the outset. All who drive know that poles could be placed further from the road, that unlit highways could be lit, that medians could be wider and so on. In short, the safety of all highways could be improved.

Leaving intuition aside, one can refute statement 'a' formally. Two kinds of argument can be used in the refutation. First, many highway design standards are *limit standards*. Limit standards do not tell the designer what the safest design is. Rather, they specify the limit of what is permissible. That is, for a certain class of highway, the radius of a horizontal curve has to be *at least X* metres long, a roadside obstacle must be *at least Y* metres from the edge of the outer lane, the gradient must be *at most Z* percent and so on. Limit standards also govern sight distances, median width, side-slopes and many other road features. Just meeting such a limit standard does not make the highway as safe as it can be; if a radius larger than X is chosen, if obstacles are placed further than Y from the travelled lane and if the highway grade is less than Z, the highway is usually safer. The same is true about wider medians, lesser side-slopes, better illumination etc. In principle then, just meeting limit standards is not a sign that the product is as-safe-as-can-be. On the contrary, just meeting limit standards it is a sign of an ungenerous design that may or may not be justified. Thus, one can certainly not claim that a road designed to meet standards is as safe as it can be. (Many elements of the new freeway the safety of which the Minister defended were actually at the limit.)

The second argument to refute statement 'a' pertains to the role of road and driver in crash causation. The implication of statement 'a' is that if standards are met and, as a result, the road is as safe as can be, then, what crashes occur on the road must be caused by drivers or by vehicles, not the road. While this belief is widespread among laymen, students of road safety believe that almost every crash is preceded by a long chain of interrelated causes, and that factors related to road design usually feature in most causal chains. This is best illustrated by the following extract:

*"Consider the following invented sequence of events. A driver was going North on an arterial road at 70 km/h where the speed limit is 50 km/h. He intended to take the loop-ramp onto Highway 407 to travel West. This ramp has a tight curve turning right. The posted advisory speed on the curve is 35 km/h. The driver apparently misperceived how much deceleration was needed to negotiate the curve and the vehicle skidded to the left. Since there is no guiderail at this point, the*

*vehicle rolled down the embankment. The embankment is a drop of 5 m over a horizontal distance of 15 m. The rear door opened, an unbelted child was ejected and severely injured.*

*Clearly the driver was ‘going too fast for the conditions’ and the child should have been wearing a seat belt. But fault or culpability are not the issue here. For us the question is: ‘what might help prevent crashes of this kind or reduce their severity?’ If we could make drivers to go slower at this site, if we could help drivers perceive what deceleration is needed, and if the curve was built less tight, then fewer vehicles would run off the road on this ramp. If a guiderail was placed along the entire curve, some skidding vehicles would be prevented from tumbling down the embankment. Even if there was no guiderail, but had the embankment been built to be less steep, some vehicles might not overturn going down the slope. If vehicles were made so that doors are less likely to open when the vehicle rolls, there would be fewer ejections of occupants. If we could induce more occupants to wear seat belts, this too would reduce the chance of ejection and injury. All these and several other actions might have altered the course of events and the final outcome.*

...

*Obviously, the driver’s ‘human error’ played a major role in this invented story, as it does in most real crashes. This leads many to think that the road users ought to be the sole target of preventive measures. Amongst road safety professionals such thinking is widely recognized as incorrect. The fact that almost all crashes could have been prevented had the involved persons acted differently does not mean that the most effective way to reduce crashes is to alter people’s behaviour or tendency to make errors. Effective action must aim jointly at the human element, the vehicle and the road. Road design can reduce the incidence of human error, road design can reduce the chance of a human error to end up as a crash, and road design can ameliorate the severity of the consequences of crashes that are initiated by human error.”(PEO, 1997, p.13,14)*

In sum, statement ‘a’, that a road meeting current standards is as safe as it can be, is indefensible.

Statement ‘b’, that roads designed to meet current standards are as safe as they *should* be, can also be refuted by several lines of argument. The first line of argument is the same as before. If a limit standard is the boundary of what is acceptable, and if safety can be improved by making the curve radius longer, the side-slope milder, the median wider and so on, then, just meeting the limit standard cannot possibly be a sign that the road is as safe as it should be. Surely one cannot claim that the appropriate level of safety is always at the boundary of what is minimally acceptable.

Second, one cannot decide how much salt to put into a soup without an anticipation of its taste. Similarly, if safety matters, one cannot come to an opinion whether to make 11 or 12-foot wide lanes standard without a fact-based anticipation of how many crashes might be saved. For a road design standard to be the embodiment of some ‘appropriate’ safety, it must be true that those who write the standards can anticipate the extent to which important road design decisions affect safety.

It may come as a surprise that, typically, writers of standards did not know how what they choose affects safety. To test the verity of my irreverent assertion is simple. One only has to ask the

highway designer or the member of the standards committee questions such as: “Approximately how many crashes will be saved by increasing the horizontal radius of this road from 100 m to 200 m; how many by making lanes 12 instead of 11 feet; or by how much will crash severity be reduced by changing this side-slope from 3:1 to 5:1?”. If they cannot answer, then the safety built into the current standards cannot be ‘appropriate’. A clear indication of the verity of my claim is the fact that till today, the waning days of the twentieth century, we do not have a tool that can predict the road safety consequences of alternative highway designs. I will return to this subject in Section 2.

The third line of argument to invalidate statement ‘b’ derives from the fact that many features of the road which impact on its future safety are not determined by standards. If so, one cannot claim that meeting standards will ensure that a road is appropriately safe. To illustrate, in the aforementioned anecdote about the new freeway that came under criticism, safety concerns were expressed mainly about the absence of a median barrier. The absence of a median barrier was of concern because of the massive unshielded posts for high-mast lighting located in the median. There are no standards that pertain to the choice of location for high mast lighting. In this case, a decision with obvious safety consequences was made without being guided by standards. Similarly, in the history of the same freeway, it became important to save on capital cost. To save cost, a decision was made to delete several interchanges from the original design. The deletion of an interchange has major safety repercussions. It will remove some of the freeway traffic to surface streets and will extend the length of some trips. More generally, the safety of any road is strongly influenced by the number of intersections, interchanges, and other access points. Yet no design standard guides this choice. Obviously, the safety of a road is influenced by many features that are not determined by adherence to standards. It follows that just adherence to standards cannot possibly bring about road designs that are appropriately safe.

In sum, neither statement ‘a’ nor statement ‘b’ are true. Roads designed to meet standards are neither ‘as safe as they can be’ nor ‘as safe as they should be’. Therefore, in no legitimate sense of the word can one maintain the roads built to standards are safe. Highway design is at present dominated by adherence to standards. There are no design tools for building a premeditated level of safety into a road. It follows that the level of safety built into roads is unpremeditated.

## **2. Of Dead dogs and other preconceived notions.**

The previous section is an arid, cerebral argument. Dry logic seldom makes a dent in strongly held beliefs. Can the amorphous process by which standards committees forge their product be judged in the light of clear-cut absolutes? Is it right to dismiss the collective judgement by many persons of good will and of considerable experience as insignificant? Must one really know by how much crash frequency and severity change as some road feature is varied before a standard is coined? Is it not sufficient to rely on common sense and experience to judge that, e.g., wider lanes or longer sight distances tend to make roads safer? To make the argument less cerebral, to diminish the belief that, by hook or crook, safety is in fact taken care off, and to lay the groundwork for the reasoning in section 3, I will present a few historical anecdotes.

The first anecdote is about the standard that pertains to the design of vertical crest curves. It shows how a preconceived idea about why crashes occur has shaped the evolution of a standard in which factual knowledge of safety was neither required nor played a discernible role.

The vertical alignment of a road is made up of straight lines connected by parabolic curves. On the straight segment the driver can see as far as vision and visibility permit. On the crest curve portion of the road, sight distance may be limited by the shape of the parabolic curve. This shape is chosen by the designer and governed by design standards. Most readers would have had the experience of driving on the upgrade of a secondary rural road where, on occasion, one eases up of the gas pedal because of uncertainty about what lies ahead. Had the designer chosen a longer and shallower parabolic curve, one could see further and the uncertainty would diminish. From the earliest time, all road design standards prescribe that the parabola be sufficiently shallow so that, if there is some object of specified height in the path of the vehicle, it can be seen from far enough for the driver to stop safely. In this manner, the standard is driven by an explicit concern for safety. The distance required for a safe stop (the 'stopping sight distance') is easily calculated from Newtonian mechanics once the speed, the grade of the road, the reaction time of the driver and the friction between the tyres and the road are given. Furthermore, if the height of the object to be seen and the height of the driver's eye above the road are given, the rest is an exercise in analytic geometry. Thus the core of the standard are the 'design speed' and a few 'parameters' ( the reaction time, pavement-tire friction, eye height and object height). The rest is a matter of computation based on physics and mathematics. The designer can compute (look up in a table) what shape of the parabola will satisfy the stopping sight distance requirement.

All this seems perfectly sensible. Note that to erect this logical edifice it was not necessary to use knowledge about how crash frequency or severity depend on the shape of the crest parabolas. All that was required was to imagine what situation on crest curves may lead to crashes. In this case the conjecture was that sight distance limitations are an important cause of crashes on crest curves.

It may be found surprising that knowledge about whether and how crash frequency on crest curves depends on the available sight distance was not needed to devise a design procedure that is driven by concern for safety. The procedure is based on a plausible conjecture. The field of road safety is littered with carcasses of plausible conjectures that did not pan out. Conjectures, no matter how plausible, are not usually acceptable when it comes to matters affecting health. Thus, e.g., a drug will not be approved for use unless its effect is carefully tested and its curative benefits as well as harmful side effects are known. Yet, the design of vertical crest curves is based not on empirical fact but on plausible conjecture. By founding road design on an unproven conjecture, the link between reality and road safety (as measured by crash frequency and severity) has been severed. The design of crest curves became a ritual founded on a preconceived idea of what causes 'failures' (i.e., crashes) to occur on crest curves.

On this an interesting story can be told (Hauer, 1988). Recall that one of the parameters in the design procedure is the height of the obstacle to be seen by the driver in time.' Originally (already in 1940) American engineering standards set the obstacle height at 4". Those who wrote this standard

did not have any particular obstacle in mind (although, rumour has it that some refer to it as the 'dead dog' criterion). They said that ". . . by increasing the height of object from 0" to 4" the required length of vertical curve is reduced by 40% . . . use of a greater object height . . . results in little additional economy . . ." (AASHO, 1954). The 'economy' here refers to earthwork, and is that of not having to dig deeper into the hill over which the road passes. Thus the 4" was selected not because lower obstacles are not a threat to safety but because the selection of a higher obstacle would not save much in construction cost. Since, at that time, nobody knew how many crashes are due to obstacles on the road, what kinds of obstacles these are, and what fraction of crashes would not have occurred had the crest been flatter, the standards committee did what was sensible. They made a decision on the basis of what was known, namely the cost of construction.

For two decades everybody was designing roads using exacting calculations to make 4"-high obstacles visible in time to stop. Then, around 1961, it became apparent that in the newer model cars the average driver's eye was much lower than a decade or two earlier. Thus, drivers of newer cars could not really see 4" objects at the prescribed stopping sight distance. Not that there was a noticeable increase in collisions with obstacles on the road; I have not found any sign that this matter has been investigated. What must have been found disconcerting was that crest curves which earlier were in accord with the standard (and therefore presumed safe) now appeared to be substandard. The solution to the predicament was not difficult. Since the 4" obstacle neither was some real object nor has it been selected on the basis of some factual relationship to safety, the Committee on Planning and Design Policies had no compunction noting that "the loss in sight distance resulting from lower eye height could be offset . . . by assuming an object higher than four inches . . ." Indeed in the 1965 AASHO Blue Book, 6" obstacles became the standard of design.

The practical men of the committee were wrestling with the surreal problem of setting the height of an imaginary obstacle of unspecified nature with which drivers collide at unknown frequency. Still, a numerical value has to be specified because such a determination is needed for the execution of a calculation which is a part of the design ritual. Under the endearing cover of an amusing anecdote are the outlines of a grave and pervasive problem: there is much concern about the rigour of form and little evidence of concern about substance.

When the Roads and Transportation Association of Canada prepared its Metric Edition of Geometric Design Standards (RTAC, 1976) a 6" object height was selected (copied) as "desirable" and a 15" obstacle height was selected as "allowable." It is to the credit of the RTAC that thought was given to the kind of object that the driver has to see; vehicle taillights which are some 15" above the road were the "obstacle" specified. In the later edition (RTAC, 1986) the 6" obstacle is not even called 'desirable' any more, it applies only to low volume roads where maintenance is sparse and where the driver may encounter logs on the road; everywhere else 15" obstacles may be used for design.

Standards are set by a 'standards committee'. Although the original motivating concern is safety, the committee usually recognizes that the relationship of sight distance on crests to safety has never been established. So the standards committee has nothing tangible to go on. But roads have to

be built and engineers are reared to be doers, not doubters. Therefore, considerations other than safety have to shape the decision. Naturally, at the end of the day a decision is made. The decision can be to use obstacles that are 0" high in Germany, 4" and later 6" in the U.S.A., 8" in Australia and 15" in Canada. Subsequently, highway designers go through the exacting ritual of designing parabolic vertical curves which comply with the current standard, however arbitrary, and do so in the deep rooted and honestly held belief that this satisfies the interest of safety.

Their designs translate into real costs. It is more expensive to build highways to ensure that all obstacles are visible and it is cheaper to build roads to ensure only the visibility of taillights. If it be true that the number and severity of crashes is not increased when only obstacles higher than 15" are visible in time to stop, why waste money on flatter crest curves? Conversely, if the number and severity of crashes does increase when only 15" instead of 6" obstacles can be seen in time, can one make a rational decision about a standard if the amount of deterioration in safety is unknown? Surely to make rational decisions of this kind requires that the relationship between sight distance and safety be known. Assumptions and conjectures based on intuition, experience and preconceived notions are insufficient.

At the time the standard for the design of crest curves came into being, little was known about safety. Today we know that only 0.07 percent of reported crashes involve objects less than 6" high (Kahl and Fambro, 1995). We also know that, till today, no link has been found between the risk of collisions with small fixed objects on crest curves and the available sight distance. On the contrary, (Fitzpatrick, Fambro, and Stoddard, 1997) say that "Crash rates on rural two-lane highways with limited stopping sight distance (at crest curves) are similar to the crash rates on all rural highways." Thus, the assumption invoked at the dawn of highway design history which allowed the formulation of a design procedure based on the avoidance of dead dogs in the middle of the road seems to have little to do with real road safety. Still, till today, the same standard stands, the same exacting but illusory constructs are used in the design of crest curves. Only the size of the dog and of other parameters is changing.

From time-to-time more research is done about what 'reaction time' or 'deceleration rate' should be used in the computation of the 'stopping distance', or about what 'driver eye height' or 'object height' should be plugged into the formula determining from how far the object can be seen. The last recommended revision makes for somewhat shallower (longer) crest curves. The authors (Fambro, Fitzpatrick, and Koppa, 1997, p. 80) rightly note that "these recommendations are based on driver capabilities and performance rather than on a need for additional safety." That is, in the absence of a proven relationship between sight distance and the frequency or severity of crashes on crest curves, the parameter values to be used have no known bearing on safety. It follows that, in spite of appearances, the design procedure for crest curves is not driven by safety but by other concerns. To cater to these 'other concerns' the design ritual may be adequate. However, since the design ritual is not based on knowledge of road safety, one may not claim that it builds into the road an appropriate amount of safety.



The second anecdote is about the width of two-lane roads. It shows that from early on, the standards committee chose to believe that the larger the separation between oncoming vehicles the better for safety. Notwithstanding the available empirical evidence counselling caution, consecutive committees stuck to this atavistic belief and wrote standards on its basis. The Committees believed that lanes less than 11 feet provide hazardously inadequate clearances between vehicles. Since many roads with 9 and 10 foot lanes existed and were being built, hard choices had to be made about minimum acceptable lane widths. Even though the essence of the tradeoff was between crashes and money, there is no evidence that in setting minimum lane width standards the committees used the then available information about what appeared to be the relationship between lane width and crash frequency.

The historical roots of the lane-width standard go back to the period of 1938-1944 when seven 'geometric design policies' were written by the Committee on Planning and Design Policies of the American Association of State Highway Officials. This group of policies was assembled into a single volume in 1950 and published with revisions as a 'Policy on Geometric Design of Rural Highways' in 1954. The 'Policy' was revised and reissued in 1965, 1984, 1990 and 1994.

About lane width the 1954 Policy says:

"No feature of a highway has a greater influence on safety and comfort of driving than the width of the surface. . . . Ten- to 12-foot lane width are now standard and the tendency is toward the larger value. . . . Observations on 2-lane two-way rural highways <sup>(1)</sup> show that hazardous conditions exist on surfaces less than 22 feet wide carrying even moderate volumes of mixed traffic and that, to permit desired clearance between commercial vehicles, a 24-foot surface is required . . . From this and similar studies it has been concluded and generally accepted that lane width of 11 feet and preferably 12 feet should be provided on modern main highways." (Pp. 192-193).

The 1954 Policy says further that:

". . . it is not economically feasible nor justifiable to utilize these standards (12 foot lanes with 10 foot shoulders) for all highways. A logical approach is to determine minimum . . . standards in relation to traffic demands . . ." (p. 223).

On this rationale, minimum lane-widths were tabulated (Policy, 1954, Table V-1). The tabulated values range from a lane width of 9 feet (when the design speed  $\leq$  50 mph and design hourly volume is 10-50 vehicles) to 12 feet (when the design speed is 70 mph and the design hourly volume is 400 or larger).

The reference in the quote <sup>(1)</sup> is to a paper by Taragin who in 1944 published an important paper summarizing extensive empirical findings about vehicle speeds and vehicle placement as a function of pavement width. Taragin was of the opinion that "The body and edge clearances for meeting vehicles or perhaps for passing vehicles are, therefore, the critical factors that determine adequate pavement width." (p. 310). In his view an "adequate pavement width" is when drivers do not shift toward the edge of the pavement when meeting an oncoming vehicle. His data indicate that

this occurs for trucks when the lanes are 12 feet wide. Based on clearances between vehicles and what drivers were seen to do, Taragin wrote in conclusion 6 that “Hazardous traffic conditions exist on pavements less than 22 ft. wide that carry even moderate volumes of mixed traffic.” (p. 317). Taragin’s conclusion has been transplanted verbatim into the quote from the 1954 Policy.

The writers of the crest curve standard assumed that the critical situation that might lead to failure (crash) was if an obstacle in the path of the driver is seen too late. The authors of the lane width standard imagined the critical situation that might lead to failure to be the loss of clearance between two oncoming vehicles. Crest curves are designed not to ‘fail’ at a specified ‘design speed’ and ‘obstacle height’. Lane widths are designed not to fail when the meeting ‘design vehicles’, i.e., trucks. Parameters for crest curve design are values selected from the distributions of some measurable properties (reaction time, pavement friction and driver eye height) to cover most but not all eventualities. The measurable properties for lane width design are separation between oncoming vehicles and how much drivers tend to shift to the right.

Taragin’s paper contains no information about crash frequency or severity as a function of lane width and yet he comes to conclusions about safety. He speculates that if drivers feel the need to shift to the right when meeting an oncoming vehicle, a hazard exists; when they no longer shift to the right, the lane-width related hazard is of no concern. Subtly, the occurrence of crashes as a manifestation of safety has been replaced by an aspect of driver behaviour. Once again conjecture is substituted for knowledge of fact. In the first anecdote the conjecture was that crashes on crest curves occur when drivers do not see obstacles in their path in time. If so, it seemed to follow by logic, that the further a driver can see the safer the road must be. In the lane-width anecdote the conjecture is that it is the loss of separation between oncoming vehicles that makes drivers shift to the right is what causes crashes. If so, since wider lanes make for more separation, it seemed to follow by common sense that wider lanes are safer. In both cases a ‘situation’ (shortage of sight distance or shift to the right and loss of separation) is substituted for ‘safety outcomes’ (crash frequency and severity). Standards are then written to govern the occurrence of ‘situations’ rather than the occurrence of safety outcomes.

The next revision of the Policy has been published in 1965. It is still based on the same single Taragin paper from 1944 and contains much the same wording as the 1954 edition (except for recognizing that now some 13 foot lanes are also being built). Amazingly, the 1984 and the 1990 edition of the Policy still rely on the same single reference from 1944 and still retain the precise same wording as the 1954 quote given earlier. The 1994 (metric) edition of the Policy drops mention of the, by now ancient, reference. Without referring to any study at all it says only that: “The wider 3.6 m (12 ft.) lane provides desired clearances between large commercial vehicles . . . ” Thus the “more clearance ergo more safety” paradigm coined in the nineteen forties continued to reign.

In sum, all the consecutive Committees in all editions of the Policy declare in the first sentence that no feature of the highway has a greater influence on safety than pavement width. Yet no edition of the Policy refers to what research says about the nature of the relationship between lane width and crash frequency or severity. Had the extant research been considered, the Committees might not have

so consistently recommended - partly on safety grounds - that 12 foot lanes be used. As I will show, important research available at the time revisions to the Policy were made indicated that 11 foot lanes are safer than 12 foot lanes. In spite of this evidence, judging by what is written, all committees must have persisted in the belief that: the wider the pavement →the larger the separation between opposing vehicles→the fewer will be crashes. To support this belief and conjecture, in the course of half a century and of five Policy revisions, all committees quoted one single study done in 1944, a study containing no evidence of the link between vehicle separation and crash occurrence. Similarly, no edition of the Policy refers to any study of how many more crashes are expected on, say, 9 foot lanes than on 10 foot lanes. Yet, somehow all committees found it possible to make the tradeoffs necessary to decide under what condition 9 foot lanes are the permissible minimum and when a minimum of 10, 11 or 12 foot lanes should be used.

Between 1953 and 1994 more than thirty research studies have been published about the relationship between safety and lane width on rural two-lane roads. My purpose here is not to review all the research findings, only to show that the writers of the Policies seemed to have been uninfluenced by a large body of empirical evidence. In particular, I will argue that doubt should have arisen about the safety benefits of using 12 foot lanes instead of 11 foot lanes.

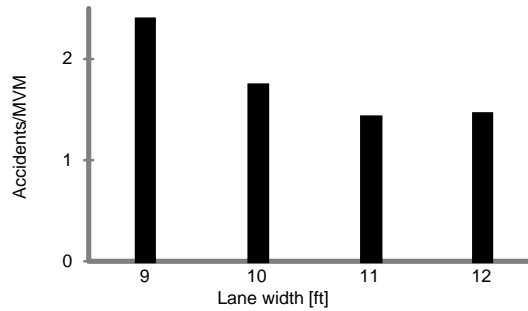
In a large-scale study with data from many states, (Raff, 1953) examined crash rates on two-lane straight roads by volume of traffic, shoulder width and pavement width. He concludes that “neither pavement width nor shoulder width nor any combination of them has a determinable effect on the crash rates on two-lane tangents.”(p.29). This finding, perhaps the best available at that time, should have raised doubt about the conjecture that crash frequency and the separation between oncoming vehicles go hand in hand. Yet, Raff’s finding did not seem to give a pause to the writers of the 1954 Policy and did not prevent them from asserting that “No feature of a highway has a greater influence on safety . . . than the width of the surface . . .” The rift between the reality of crash occurrence and the conjectured situational cause was already complete. Conjecture, not empirical findings, carried the day.

One of the classical studies is by (Belmont, 1954). Table 1 is based my re-analysis of Belmont’s data. It shows that under identical traffic conditions, roads with 10 foot lanes have 5% more crashes and roads with 12 foot lanes have 1% more crashes than roads with 11 foot lanes.

**Table 1.** Crash modification factors based on Belmont’s data.

Pavement Width in ft	18	20	22	24	26	28	30
CMF	1.21	1.05	1.00	1.01	1.06	1.13	1.21

The same ‘bottoming out’ emerges from a study of data from Louisiana by (Dart and Mann, 1970) who show the relationship in Figure 1.



**Figure 1**

Perhaps the best known is a study by Roy Jorgensen Associates, Inc.(1978) who found the results in Table 2.

**Table 2.** Initial Crash Modification Factors (p.18).

Shoulder width [ft]	Pavement width [ft]			
	≤ 18	19-20	21-22	≥ 23
1-2	1.76	1.55	1.41	1.57
3-4	1.44	1.27	1.15	1.29
5-6	1.27	1.13	1.02	1.14
7-8	1.14	1.00	0.91	1.02
>9	1.11	0.99	0.90	1.00

The authors noted that the increase from the 21-22 ft pavement width category to the ≥23 ft group is “inconsistent with the expectation” but that “it is consistent with the research findings”(p. 20.) Sacrificing ‘research findings’ to ‘expectations’ they decided to join the two rightmost columns into one pavement width category. The decision to use one width category for all pavements wider than 21 feet avoided the appearance of conflict between results of research and the words in the Policy. McLean (1980, p.192) questions this decision and maintains that while the increase in crash rates from the 21-22 ft category to the >23 ft category “. . . may have been anomalous in terms of conventional engineering expectations, they are consistent with the general hypothesis of an interaction between driver behaviour and geometric standard.”. The questionable decision to merge columns resulted in the oft-quoted accident crash modification factors (from their Table 13) reproduced in row 2 Table 3 as CMF (merged columns). Were the unmodified data used the result would be that row 3 shown as CMF (original results).

**Table 3. Modified CMFs**

1	Pavement Width in feet	≤ 18	20	22	24
2	CMF (merged columns)	1.18	1.04	1.00	1.00
3	CMF (original results)	1.25	1.10	1.00	1.11

Note that the original results are a more pronounced version of the findings based on Belmont’s data and indicate that for rural two-lane roads, making pavements wider than 22 ft was at that time detrimental to safety. A few years later Zegeer et al. (Zegeer, Deen, and Mayes, 1980; Zegeer, Deen, and Mayes, 1981) also find an upturn in the crash rate after a lane-width of about 11 feet.

I am not convinced that if research was done on current data, 12 foot lanes would be found less safe than 11 foot lanes. Much has changed since then; trucks grew to be larger and research methods have improved. However, at the time the Policy was being written and rewritten, the aforementioned findings by respected researchers should have sounded the alarm. Specifically, the interest of road safety was apparently inconsistent with the oft-repeated recommendation that lane widths of “preferably 12 feet should be provided on modern main highways.” More generally, since 12 foot lanes were apparently less safe than 11 foot lanes, the paradigmatic conjecture that more separation between oncoming vehicles means more safety should have been questioned and deposed. But no alarm was to be heard and no change to the governing paradigm was made. Safety continued to be in the domain of conjecture and common sense judgment and empirical fact was not allowed to intrude on these.

It is impossible to know what members of the Committees knew about the relationship between lane-width and crash frequency. Going by what they wrote implies that till 1994 they relied only on Taragin’s 1944 conjecture. The judgement that the Committee members had to make (about what lane width is justified in what conditions) is not an easy one. Arguments of cost, capacity, safety and comfort must be considered, and it is not clear how far quantitative cost-benefit calculations can be pushed. It is clear, however, that the safety portion of the argument should be based on crash frequency and severity. If the safety portion is based on the conjecture about separation between oncoming vehicles, and since the relationship between separation and safety is unknown, safety is not really being taken into account and the resulting standard builds an unpremeditated amount of safety into roads.

The third story is about horizontal curves. It shows with clarity the prototypical paradigm guiding the minds of writers of geometric design standards:

1-assume how failure arises ⇒

2-use physical sciences and mathematics to represent the failure situation ⇒

3-postulate ‘design loads’ and choose ‘conservative’ values for parameters ⇒

#### 4-compute values for design.

At first sight, the assumed mode of failure on which the design of a horizontal curve is based is patently logical. To move around a curve, the vehicle must be pushed by a sufficient external force acting toward the centre of curvature. If the available force is insufficient, the vehicle will drift to the outside of the curve and leave the road. This is thought to constitute ‘failure’ in this case. The faster the vehicle travels, the larger the required force. Conversely, the larger the radius of curvature the lesser is the required force. The required force is provided partly by the tire-road friction and partly by the banking of the road (the ‘superelevation’). For the assumed mode of failure (drifting out of the curve due to insufficient centripetal force), the laws of physics specify the relationship between speed, radius, superelevation and side friction. These laws can be captured by a simple mathematical formula. The formula is then used for design. Specifically, using the maximum allowed superelevation and a ‘conservative’ value for the side friction for various design speeds, one can compute the smallest “safe radius”(Policy, 1984, page 174). The design speed is the ‘design load’ in the paradigm, the maximum superelevation and the selected side friction factor are the ‘parameters’.

Interestingly, the ‘conservative’ value for side friction is not what might be encountered on “pavements that are glazed or bleeding . . . because these conditions are avoidable and geometric design should be based on acceptable surface conditions” (Policy, 1984, pp.165-166). Rather, it is based on the observed driver behaviour and derived from tests conducted about the amount of side friction that drivers will accept without slowing down when going round curves at, what they think are, safe speeds. These friction factors are conservative because they are still believed to “provide ample margin of safety against skidding” (p. 166).

Since the logic seems impeccable and conservative parameter values were used, failures should be rare. If so, one might justifiably expect that curvature has been adequately accounted for by design, and therefore, that the presence of curvature should not materially increase the chance of crash occurrence. It would then follow, that the safety of horizontal curves should not be much different from the safety of straight road sections. But this is factually grossly untrue. Ample data show that on horizontal curves crashes are much more frequent (perhaps by a factor of 3, on the average) than on straight road sections. Also, numerous studies show that the shorter the radius of a curve the higher its crash rate. In fact, what is perversely called the smallest ‘safe radius’ is the radius associated with the highest crash rate. How was the intent to design safe curves so grotesquely thwarted? How did the subversion of language arise in which what is called the ‘safe radius’ means least safe radius?

The immediate cause of the tragic disjunction is that, as in the earlier anecdotes about vertical curves and lane width, no empirical information about crash occurrence has been used to develop the design procedure for horizontal curves. Nor did anyone seem to consider how crash frequency and severity depend on curve radius or superelevation. The feat of designing for safety without using the extant empirical knowledge of safety was facilitated by the apparent legitimacy of the aforementioned design paradigm.

First, it was assumed to be obvious that failure results when there is insufficient force to keep an object moving at the design speed on a circular path. In this wholly mechanistic conception there seems to be no room for the driver who has to actually steer the vehicle on the curved path at an appropriate speed. In fact a large proportion of crashes on curves occur when the driver did not anticipate the curve correctly and did not follow the curve of the road. Belated reaction often results in over-correction and loss of control. In these instances the availability of an adequate centripetal force on an assumed circular path has no relevance and little influence. Were the mechanistic conception the main mode of failure true, vehicles should be running off the road only on the outside of curves. However, data show that 11%-56% of the vehicles run off on the other side of the road (Bissell, Pilkington, Mason, and Woods 1982). So that there is a substantive discord between how failures are assumed to occur and between the reality of crash occurrence.

Second, the role of the 'design load' in the general paradigm is played here by the 'design speed'. Normally, design loads are so selected that their probability of being exceeded is sufficiently small. Only then can failure be appropriately rare. But the design speed used in geometric design standards has only the vaguest relationship to any real rarity of occurrence.

The design speed is defined somewhat circuitously<sup>1</sup> as "the maximum safe speed that can be maintained over a specified section of highway . . ." (Policy, 1984, page 60). However, in fact, the speed at which drivers negotiate curves routinely exceeds the design speed. Krammes (1994) reports that the 85% percentile speed exceeded the design speed on the large majority (about 90%) of curves where measurements were made. Similar findings for Australia were published by McLean (1981) That is, much more than 15% of the drivers traverse curves at speeds that are larger than what has been assumed for design. This is certainly nothing rare. Naturally, the driver can have no knowledge of the 'design speed' that has been used in the designer's calculations. Since the design speed has no clear relationship to either the speed limit or the speed expected to be exceeded by only a very small proportion of drivers, it is entirely unclear what it represents or why it ought to be relevant to curve design.

Having mentioned two faults of how the general design paradigm has been interpreted in this case, I could go on and question the soundness of using sensations of comfort as an acceptable determinant of the proper margin of safety for side friction rather than values very rarely found on actual pavements. But this seems hardly worthwhile. It is by now clear that there is no premeditated connection between the reality of crash occurrence on horizontal curves and the procedure used for their design.

---

<sup>1</sup> In plain language the definition means something like: "If you will traverse the curve at the speed not exceeding the speed for which I have designed it, you can steer the prescribed path without feeling the need to slow down and under most circumstances there will be enough force to keep you on that path". In short, the design speed is the speed chosen for design. This is what makes the definition circuitous.

I have recounted three anecdotes that illustrate the rift between the intent to build appropriately safe roads and the guidance provided by the succession of policies on geometric design. The implied criticism is perhaps too harsh if three important aspects of reality are not remembered. First, research tends to produce diverse results. This is true in all fields of inquiry. When for a study claiming one thing there is another study claiming the opposite, then, amidst practitioners, this will be taken as license to disregard research results altogether. Second, the validity of research results is often easy to question. In road safety we do not have the luxury of randomized experiments that allow clear interpretation. The road safety researcher attempts to interpret ‘haphazard data’, data that happen to be available and come from a world where many things change simultaneously and are interrelated. Such data, as a rule, do not lead to demonstrably valid conclusions. Such unconvincing and contradictory research results invite dismissal and thus legitimize reliance on judgement, common sense, and preconceived notions. The freedom of action provided by the legitimization of judgement unhampered by empirical fact is important in view of the third element of reality - litigation. Liability is often construed in the light of documents such as the Policy. If so, what goes into the Policy is written with the litigation lawyer in mind.

These three aspects of reality perhaps explain the circumstances in which the rift between the intent and action evolved. It is not a sufficient excuse. A road is a man-made product. In use, it is known to be harmful to health. It is not acceptable to produce roads and to put them into use without providing for a premeditated amount of safety. To suggest remedies to what is deemed unacceptable, it is important to seek the roots of this rift. This is the subject of the next section.

### **3. The burden of history.**

Highway design is a collection of decisions: what will be the shape of a crest curve, how wide to make the lanes, what should be the radius of a horizontal curve and so on. These decisions affect to varying degrees the frequency and severity of future crashes. The common thread of the three anecdotes told earlier is that, notwithstanding the declared primacy of concern for safety, highway design standards, policies and procedures that guide design decisions tend to be formulated without the knowledge of how the design decisions are likely to affect future crashes. It is difficult to understand how this sorry mess came about and why the sincerity of intent has been misdirected. One strand of explanation is in the historical evolution of civil engineering, the cradle into which most highway designers are born, the culture into which they are socialized.

Many modern road design practices bear the imprint of the earlier railroad building era. From that period of history we have inherited spiral transitions, the concept that horizontal curves are designed for certain speeds, the need for superelevation, and the general mode of thought that goes with all these. This continuity of tradition can be traced further back to the common ancestry of civil engineering and perhaps of all engineering. Engineers tend to base design procedures on the foundation of physical laws, mathematics and the empirical knowledge of the properties of materials. Given some ‘design targets’ that often have only an intuitive and practical justification (design loads, design winds, design storms, design throughput etc.) one strives to make ‘failure’ adequately rare.



Elements of this tradition are evident in all three cases reviewed earlier and have been explicitly noted. There are two major problems with this heritage when it is applied to safety in road design.

The first problem is the persistent tendency to define failure by surrogates, rather than to define it in terms of expected crash frequency and severity. A beam fails when it cracks, collapses or unduly deflects; a culvert fails when water overflows and damages the surrounding; the pavement fails when it is full of potholes, heavily rutted, cracked etc. In all these cases the concept of failure is defined by events more-or-less self-evidently linked to the harm of failure. This is not so for safety in highway design. The real harm of failure are crashes and their consequences. However, the tendency is to supplant crashes by ostensibly related surrogates, such as a shortfall in sight distance, lack of clearance between oncoming vehicles, or insufficient centripetal force.

To the members of the early standard committees the link between an insufficient sight distance and crashes on crest curves may have appeared self-evident. It may also have seemed self-evident that the wider the lane the safer the road, or that vehicles run off curves because of insufficient centripetal force. The working assumption was that these surrogates have a clear link to crashes.

In the event, so far, researchers have not found a link between sight distance and crash frequency on crest curves. Therefore, even if such a link exists, it likely to be weak. A weak or nonexistent link is an insufficient foundation for a major design procedure. If substandard sight distances on crest curves are not associated with an increase in crash frequency and if generous sight distances do not seem to save crashes, then this surrogate concept of failure may not have been well chosen.

In the case of lane width, the surrogate for crashes was the separation between oncoming trucks and the extent to which drivers felt the need to shift to the right when meeting another vehicle. However, at the time most research on this subject was done, the extant evidence seemed to indicate that in spite of the wider separation and the lesser perceived need to shift, 12 foot lanes on rural two-lane roads were less safe than 11 foot lanes. If so, once more, the surrogate for crashes was ill chosen. It may have wasted both money and lives.

Similarly, members of the early standard committees imagined or assumed that failure on a horizontal curve occurs when pavement friction and superelevation are insufficient to provide the force needed to turn the vehicle around the curve and, as a result, the vehicle 'slides out'. Whether this is the main mechanism by which crashes occur on curves is in doubt. We know that the resting position of vehicles for a large proportion of crashes is on the inside of a curve. These could not have 'slid out' and therefore were not caused by the mode of failure imagined by the standard and the design procedure. It is also clear that many (perhaps most) curve crashes occur at speeds under which side friction + superelevation were sufficient to keep the vehicle moving along the curve. The driver may have mis-perceived the conditions, may have been impaired, tired or distracted and did not steer along the prescribed path. In this case too, the design process is based on an imagined or

assumed concept of failure that may be linked to but one mechanism of crash occurrence while leaving out several others.

The attraction of using surrogates is twofold. First, in a situation in which it is not known what are the crash frequency and severity consequences of a decision, and when a decision has to be made, the use of judgement to define sensible surrogates is fully justified. Early standards committees must have found themselves in this circumstance. Second, if a clear causal link has been established between crashes and some surrogate, it is often best to observe or predict changes in the surrogate as a stepping stone for estimating the change in crash frequency and severity. Thus, e.g., there is considerable consensus about how speed affects crash severity. Therefore, if some intervention is expected to affect speed, and when later speed change has been observed, one may make claims about the corresponding changes in crash severity. When, eventually the changes in crash severity are estimated and confirmed, the knowledge and understanding of the causal mechanism that is provided by the surrogates adds confidence in the result.

The danger of supplanting the real measure of safety (i.e., crash frequency and severity) by surrogates arises when the link between the two is conjectural, when the link remains unproven for long, and when the use of the unproven surrogates becomes so habitual that the need to eventually speak in terms of crashes is forgotten. This danger would not be great if intuition about the surrogates being good indicators of crashes usually panned out. Unfortunately, in road safety, intuition is a fallible guide and plausible conjectures often turn out to be incorrect. The three anecdotes told in Section 3 illustrate this fallibility. We also know of many interventions that by common sense should have worked and later were found wanting. Suffice to mention driver education in high schools, ABS for vehicles, painted crosswalks for pedestrians, resurfacing for rural roads etc. By intuition one expected that providing professional instruction for skills and attitudes, making for more control in braking, warning drivers of a reserve for pedestrians, or providing a new pavement with more friction, that all these should reduce crashes. In fact some of these measures were useless and some seem harmful. After the fact we can find excuses: speed adaptation, false sense of security, etc. Making these excuses means that road user responds to change in many ways, that some responses are not observable, and that, we are not yet clever enough to say how the bottom-line will change as a result.

It is possible that the writers of standards at all times invented these various surrogates of failure more out of concern for legal or moral liability than out of a desire to strike a balance between safety and cost. Thus, e.g., for legal and moral reasons, the sight distance should be long enough for drivers to stop safely, the superelevation should be large enough to hold the vehicle on the curve at legal speeds, the intergreen period at a signal should be so that the driver can either stop safely before the stop line or enter the intersection before the onset of red. If roads are built in accord with such principles, then professional decisions are easier to defend. If this is the case, then standards are not the guardians of safety, they guard against liability.

In sum, much of the geometric highway design is ostensibly motivated by safety concerns. However, its link to safety is often 'second hand', conjectural and, at times, illusory. It ought to be obvious that the concept of 'safety-related failure' in road design must be rooted in crash frequency

and severity. Safety failure is not a matter of ‘either - or’, but a matter of degree. It is not like the collapse of a roof or the flooding of a culvert, but more like the deflection of a beam exceeding the allowable amount or the cracking of a pavement occurring prematurely. Accordingly safety failure should be defined straightforwardly and directly in terms the expected frequency of crashes or crash consequences. Surrogates can be used only if they have a known relationship to crash frequency and severity.

The second problem with the civil engineering tradition of standard setting in road design is also fundamental. Civil engineers (highway designers) are trained to deal with inanimate matter. We deal with loads, flows, modulus of elasticity, stress, strain, porosity etc. Therefore, once we understand the physics of the situation and know the properties of the materials, we can fairly well predict ‘what will happen if’. This is the basis on which reasoned design choices are made. In geometric design one central circumstance is different. Roads are built for road users. Unlike inanimate matter, road users adapt to the prevailing situation. The next time you drive on a short crest curve, pay attention to how you ease the pressure on the gas pedal, perhaps even brake slightly when the available sight distance is short. Surely you approach a tight horizontal curve very differently than you drive around a large-radius curve. Thus, in geometric design, one should not assume that speed, reaction time and similar ‘design parameters’ are quantities that do not depend on the design itself. There is no parallel to this in other civil engineering design. One does not assume that the load will be adapted to the strength of the beam or that it will rain less if the diameter of a culvert is small.

It must have been tempting for the early thinkers about geometric design to cast the road user into the familiar mode of parameters drawn from a distribution representing some invariant properties. After all, this was so successful in the characterization of concrete by compressive strength and of steel by Young’s modulus. By doing so they have erected a conceptual framework which cannot recognize the basic fact that people adapt to circumstances whereas inanimate matter does not. For road safety it is a flawed and deficient framework. The consequence of this fundamental misconception is that speed, reaction time and similar parameters are treated as constants in all the formulae and computation that are at the root of geometric design standards.

The thrust of this argument is that while highway design standards are perhaps motivated by concern for safety, they are the guardians of safety in a very limited sense. First, because an early tradition has been established that links design to assumed or imagined surrogate modes of failure, rather than to crash frequency and severity. Second, because highway design standards still treat the world as if it could be adequately described through physics and properties of materials and fail to recognize that road users adapt their behaviour to the road they see and expect, that what road users do depends on what the designer puts in front of them.

#### 4. What to do?

I have argued at length that the guidance of highway geometric design standards is not based on the relationship between highway design decisions and their safety consequences. As a result, the level of safety now built into roads is unpremeditated.

There are powerful reasons for not acknowledging and spelling out the link between road design and safety. The most compelling reason is the need to protect state and municipal governments against the financial perils of liability. Such perils are minimized if one can hide behind the tautological phrase that: “The road is safe because it has been built to standards”, especially since the standards are written by employees state highway agencies and issued by the American Association of State Highway and Transportation Officials.

On the other hand, there are powerful reasons for insisting on a reform of the highway design process so that it be knowledge-based and safety conscious. The main reason is that roads are man-made products that affect human safety. Users of roads have no choice but to travel on what is produced for them by others. Therefore there is an implicit relationship of trust between the travelling public and the agencies and professionals in their employ who produce roads. The essence of the trust is that road users may rightly expect the agency and its employees to use the best available knowledge in order to decide how much safety to build into the roads they produce. Not using such knowledge amounts to a breach of trust. This should be no more acceptable than the unregulated marketing of drugs or toys known to be dangerous.

Perhaps the way out is by creating a clear distinction between two kinds of safety:

- *Substantive safety* is the expected crash frequency and severity;
- *Nominal safety* is examined in reference to compliance with standards, warrants, guidelines and sanctioned design procedures.

Substantive safety is a matter of degree. A road in use cannot be safe, only safer or less so. What level of substantive safety is appropriate is therefore governed by considering what is level of safety is attainable with resources available. In contrast, a road can be nominally safe, meaning that it conforms to standards, warrants, guidelines and sanctioned design procedures. Whether a road that is nominally safe is always (or even usually) substantively safer than a road that is not nominally safe cannot be said. For example, if the standard calls for 3.75 m lanes then neither roads with 3.00 m lanes nor roads with 3.60 m lanes are nominally safe. But, in terms of substantive safety, roads with 3.00 m lanes are known to have many more crashes than roads with 3.75 m lanes while the same cannot be said about roads with 3.60 m lanes. For all we know 3.60 m lanes may be as safe or safer than roads with 3.75 m lanes. In short, nominal safety and substantive safety are two distinct aspects of a road, one is a ‘yes’ or ‘no’ determination, while the other is measured on a continuous scale, and the two may or may not go hand-in-hand, as the case may be.

If the trust between road users and road producers is to be maintained, consideration of substantive safety in road design is obviously important. The question is whether there is something important about nominal safety that is worth preserving. Four aspects of nominal safety have value.

- Our designs must enable road users to behave legally. This can be attained by nominal safety.
- Our designs should not create situations with which a significant minority of road users has difficulties. This too can be secured by making roads nominally safe.
- Nominal safety is useful protection against claims of moral, professional and legal liability.
- Resorting to nominal safety may be a temporary necessity when crash frequency and severity consequences are unknown. In such cases, a statement about the absence of crash-based information is needed.

The ability of road users to behave legally is an important consideration and it is different from concern about substantive safety. Thus, e.g., if it is illegal to enter a signalized intersection on 'Red', the designer must ask what duration of amber will allow almost all road users who decide to proceed at the onset of amber, to enter before the onset of red. In contrast, when one is concerned about substantive safety one asks what duration of amber will be associated with fewer crashes. The answers to these two questions may or may not be the same. (The ability to behave legally is not always of primary importance. Recall, e.g., that at a rail-highway grade crossing the red indication is not preceded by amber. Here the concern about the ability to behave legally seems to be secondary to the ability to be clear about who is to blame for a collision.)

Concern about situations which make the use of the road difficult for some road users is also an important consideration. Thus, e.g., even slow walkers should be able to reach the opposite curb during the time allotted to a protected pedestrian crossings. In this case one asks how long should be the WALK signal so as to serve a certain high proportion of the population walking speeds (often the mystical 85 percentile.). The answer to this question may be entirely different from the answer to the question what duration of WALK makes for fewer crashes. First there is no evidence that long durations of 'WALK' make for fewer pedestrian crashes. Second, inasmuch as longer 'WALK' gives less green to cars, it makes for more delay, longer queues, more stopping and perhaps more car crashes.

The third aspect of nominal safety, the issue of liability, deserves careful thought. Lawyers tend to judge the adequacy of a design or a road with reference to what is accepted professional practice; their demarcation lines between what is acceptable and what is substandard is usually sharp. A shoulder width may be deemed substandard even if it is only a few inches narrower than what the applicable standards specify. In contrast, transportation professionals, those who write standards and design by them are cognizant of the limitations of knowledge and know how large is the role of judgement in the formulation of standards. They view the world in shades of grey; they think of cost, effect and takeoffs. But cost and effect are weak arguments in the court of law. For this reason the safe haven offered by nominal safety and design-by-standards is inviting. One can always determine without equivocation whether a road or a design is nominally safe. This has a corroding effect on engineering practice. Rather the designing for what is appropriate, design is often for what is

defensible. Since what is sanctioned by standards is in many cases the minimally acceptable, there is pressure to produce the cheapest and minimally acceptable roads. In any case, since defence against liability is a fact of life, the standards and warrants based on nominal safety will also remain a fact of life.

Finally, there will always be design decisions the safety consequences of which are not yet known. If guidance has to be provided, it can come only from the accumulated understanding of crash causation and human behaviour. Such understanding can be the basis of *provisional* standards, warrants, guidelines and sanctioned design procedures. Because our current understanding of crash causation and human behaviour is imperfect, our anticipation of safety consequences has often been wrong. Therefore, concerted and prompt action must always be taken to confirm the validity of speculation. The confirmation is always by safety outcomes; that is, by crash frequency and severity consequences. Only such a confirmation can remove the label ‘provisional’.

I am led to conclude that there are certain aspects of nominal safety and its embodiment in standards, warrants, guidelines and sanctioned design procedures that will be preserved and some that are worth preserving. This does not mean that no change in the standards is required. Two fundamental shortcomings of the paradigm on which geometric design standards are based have been identified earlier. First, that there is a tendency to define failure by surrogates, rather than to define it in terms of expected crash frequency and severity. Second, that road users are dealt with as inanimate matter, as if it was possible to represent them by fixed parameters that do not depend on the design, as if they did not adapt to the road which the designer puts in front of them. A revision of this design paradigm is in order. In addition, it would help if up-to-date information about the safety consequences of design decisions was made an explicit part of the various Policies and Guides on geometric design.

Having dealt with nominal safety and its reform I now turn to substantive safety. The question is how to reform the highway design process so that an appropriate amount of safety is designed into roads. Since this is not presently done, the answer is obvious: make the consideration of substantive safety an explicit and knowledge-based part of the highway design process. While the answer is straightforward, implementation is not easy. To succeed three elements must be in place:

- a. That the best current knowledge about the relationship between highway design decisions and their safety consequences be readily available to the designer;
- b. That those designing roads be trained (certified) in road safety and educated about the relationship between highway design and safety;
- c. That political-public guidance be given to highway designers on what level of safety to aim for.

The store of existing knowledge about the relationship between specific highway design decisions and their safety consequences is rich but difficult to tap into. It is dispersed in reports, journal papers, libraries, and in people’s heads. Also, it is continuously evolving. It takes time and

considerable expertise to assemble, sift through, and assess, what has been published over the years on a certain subject. The highway designer cannot do so. Accordingly,

- reform of the highway design process requires the periodic issuance of an authoritative document summarizing what is known about the safety repercussions of highway design decisions.

Having taught undergraduate civil engineers for decades, I know that in the course of their undergraduate training they get about one term of exposure to the rituals of nominal safety and not a single lecture about substantive safety. They become highway designers without any idea of how their designs are likely to influence future crash frequency and severity. I cannot think of any other field of professional endeavour in which this is permitted. Accordingly,

- reform of the highway design process requires that those who sign highway design documents be certified to have been appropriately educated in highway safety.

No matter how strong the desire to skirt the issue, highway design choices involve a takeoff between resources and life or limb. It is always possible to save life and limb by making the median wider, by installing illumination, by removing trees and poles from the roadside, etc. There is nothing in the education or the status of the highway design engineer that permits her to judge what level of safety is appropriate. This is a judgement that has to be made by road users and their representatives. Accordingly,

- reform of the highway design process requires that political-public guidance be given to highway designers on what level of safety to aim for.

## 5. Summary.

I have argued that roads designed to standards are neither safe nor appropriately safe. This I tried to show both by logical argument and by historical examples. Highway design standards have evolved within the straightjacket of a design paradigm that is deficient for the purposes of road safety. They are based on surrogate concepts of failure, not on the frequency or severity of crashes. Furthermore, current design standards try to represent road users by certain fixed parameters and fail to recognize the fact the road user remembers the roads travelled and the road behind and adapts to the road seen ahead. As a result, the relationship between highway design standards and road safety is unclear and the level of safety designed into roads is unpremeditated.

Reform of the highway design process requires the recognition of separation between two distinct concepts of safety. *Nominal safety* is judged by compliance with standards, warrants, policies and sanctioned procedures. It ensures that most road users can behave legally, that design does not make road use difficult for significant minorities and provides protection from moral, professional and legal liability. To reform how nominal safety is dealt with, the faulty design paradigm has to be

replaced by a new one, and genuine safety information should be incorporated in highway design standards. The concept of *substantive safety* is measured by expected crash frequency and severity. Strange as it sounds, substantive safety is a new concept to be introduced into highway design process. The introduction of substantive safety into highway design requires three action elements: that what is known about the relationship between safety and design decisions be authoritatively summarized and periodically reissued, that those who sign design documents be certified to have mastered the currently available knowledge, and that political guidance be provided to designers on what level of safety to design into roads.

Finally, an apology is due. Therefore there is an element of the unfair in my focus on the succession of committees that wrote the geometric design standards for AASHO and later AASHTO. After all, there are many standards other than those for geometric design that have only a tenuous link to safety. Thus, e.g., it seems acceptable to use medical opinion as a sufficient basis of the static visual acuity requirements for driver licensing, even though its correlation to crash experience is weak to nonexistent. My excuse is that I have drawn my examples from what I am familiar with and I wrote about what concerns me - the role of civil engineers in the delivery of road safety. ("Only you did I know of all the families of the earth, this is why on you I will visit all your sins." Amos, 3, 2). There was no intent to be critical of either persons or organizations that acted in the same way as many others do.

#### References.

- AASHO. (1954) A policy on geometric design of rural highways. American Association of State Highway Officials, General Offices, Washington, D.C.
- PEO. (1997) Highway 407 safety review. Professional Engineers Ontario, Toronto.
- Belmont, D.M., (1954), Effect of shoulder width on crashes on two-lane tangents. *Highway Research Bulletin* **91**, Washington, D.C., 29-32.
- Bissell, H. H., Pilkington, G. B., Mason, J. M., and Woods, D. L., (1982). Synthesis of safety research related to traffic control and roadway elements. Chapter 1. FHWA-TS-82-232. Federal Highway Administration, Washington, D.C..
- Dart, K. O., Mann, L. (1970), Relationship of rural highway geometry to accident rates in Louisiana. *Highway Research Record* **313**, Washington, D.C. 1-15.
- Fambro, D. B., Fitzpatrick, K., and Koppa, R. J. (1997). Determination of stopping sight distances. Report 400, National Cooperative Highway Research Program, AASHTO and FHWA, Washington, D.C.
- Fitzpatrick, K., Fambro, D. B., and Stoddard, A. M. (1997) Safety effects of limited stopping sight distance on crest vertical curves. Paper presented at the 76<sup>th</sup> annual meeting of the Transportation Research Board.
- Hauer, E. (1988). A case for science-based road safety design and management. A case for science-based road safety design and management. In: Stammer R.E., (ed.) Highway Safety: At the crossroads. American Society of Civil Engineers.



- Kahl, K. and Fambro, D. B.,(1995), Investigation of object-related accidents affecting stopping sight distances. *Transportation Research Record* **1500**, Washington, D.C., 25-30.
- Krammes, R. (1994) Design speed and operating speed in rural highway alignment design. Annual meeting of the Transportation Research Board.
- McLean, J.,(1981), Driver speed behaviour and rural road alignment design. *Traffic Engineering and Control*. **22** (4), 208-210.
- McLean, J. R. (1980) The safety implications of geometric standards. Canberra.
- RTAC, Manual for geometric design for Canadian roads. Metric edition. Roads and Transportation Association of Canada. Ottawa.
- Raff, M. S.,(1953), Interstate highway-accident study. *Highway Research Bulletin* **74**, Washington, D.C. 18-43.
- Roy Jorgensen Associates, Inc.,(1978). Cost and safety effectiveness of highway design elements. Report 197. National Cooperative Highway Research Program, Washington D.C.
- Taragin, A. (1944) Effect of roadway width on traffic operations - two lane concrete roads. Proceedings of the 24<sup>th</sup> annual meeting of the Highway Research Board. Washington, D.C.
- Zegeer, C. V., Deen, R. C., and Mayes, J. G.,(1980). Effect of lane width and shoulder widths on accident reduction on rural, two-lane roads. Research Report, Kentucky Department of Transportation.
- Zegeer, C. V., Deen, R. C., and Mayes, J. G.,(1981), Effect of lane width and shoulder widths on accident reduction on rural, two-lane roads. *Transportation Research Record* **806**, Washington, D.C., 33-43