ICORSI

Independent Council for Road Safety International

International Symposium

Road Safety Around the World: Future Concerns

19 March 2018

Safety research and practice: problems of coexistence

Ezra Hauer

THE UNIVERSITY OF CHICAGO CENTER IN PARIS

Acknowledgement

ICoRSI has received support from:

Institutional donors: Tata Education & Development Trust, Mawana Sugars, University of Chicago Center in Paris, Veolia

Recommended citation

Hauer, E. (2018) Safety research and practice: problems of coexistence. Presented at the ICoRSI International Symposium *Road Safety Around the World: Future Concerns,* 19 March 2018, Paris . Independent Council for Road Safety International*, www.icorsi.org/icorsi-pblications.*

Declaration

The presentations included in this volume were prepared for discussion at the ICoRSI International Symposium *Road Safety Around the World: Future Concerns,* 19 March 2018, Paris. The papers may be published subsequently in a different form.

Contents may be reproduced with attribution to authors.

© Independent Council for Road Safety International

CONTENTS

1 Introduction 2

2 The First Anecdote: Two studies about the safety effect of edgelining 3

2.1 Why study the safety effect? 4

2.1.1 Issue 1: Two purposes of evaluation research 5

2.2 Between opinion and evidence 5

2.2.1 Issue 2: The road is a product 5

2.3 Opinion is a fallible guide 6

2.3.1 Issue 3: In road safety opinion-based anticipation is insufficient 7

2.4 Road users have rights 7

2.4.1 Issue 4: The right of road users 8

2.5 The shackles of dependence 8

2.5.1 Issue 5: The importance of research independence 10

2.6 An interim summary 10

3 The Second Anecdote: Rulemaking about pavement marking retroreflectivity 10

3.1 The 'Synthesis' in the service of rulemaking 12

3.2 First set of issues: Obstacles to learning from experience 15

3.2.1 Issue 6: Separation of profession-serving evaluation research from execution 16

3.2.2 Issue 7: Meta-analyses exaggerate safety benefits 16

3.2.3 Issue 8: Systematic reviews, not just meta-analyses 18

3.2.4 Issue 9: No plan, no framework 19

3.3 The second set of issues: Disregard of research findings 20

3.3.1 Issue 10: An Inquiry is needed 21

3.4 The third set of issues: Perils of the funding-finding loop 21

3.4.1 Unfounded beliefs 21

3.4.2 Issue 11: Guarding against unfounded beliefs 22

3.4.3 Harm to trust 23

3.4.4 Issue 12: Insulate research from extraneous interests 24

4 Closure 24

REFERENCES 24

Appendix A. The 'Synthesis' on the safety effect of edgelining 29

Appendix B. The 'Synthesis' on speed 35

Appendix C. The 'Synthesis' comments on Bahar et al. 37

Road safety research and practice; problems of coexistence

Ezra Hauer, Revision 2, 1/29/2018

This is a work in progress. Am I on the right track or is my perception of reality lacking in balance? Please send comments to ezra.hauer@utoronto.ca and feel free to disseminate.

Abstract. What is the relationship between road safety research and the practice by which the road infrastructure is built and operated? The question is seldom asked. I discuss the complexities of the research-practice symbiosis in the light of two historical anecdotes. These allow me to point out twelve issues of concern. My general conclusion is that the relationship as it evolved over time is unpremeditated and occasionally dysfunctional. Findings can be easily disregarded and questionable results given a ring of consensual truth. Practice based on such knowledge cannot be evidence-based. In the interest of road-user safety it is time to endow the research-practice relationship with a clear and purposeful structure.

1 INTRODUCTION

The role of applied research is to produce applicable information which, if used, makes for better practice. By 'practice' I mean all the decisions that affect road-user safety and go into the production and operation of the road network: the standards, the warrants, the laws, the campaigns, the countermeasures, etc.

In recent years there has been a veritable explosion of road safety research; more conferences, more papers, more journals. How does all this activity affect practice? How much of the research we produce is applicable? To what extent is the applicable information being used? What are research findings used for? These questions are seldom asked.

I will argue that, in North America, the relationship between road safety research and practice was and remains unstructured; that the gossamer weave that currently links the two is by and large unpremeditated; I find that research findings are easy to disregard, and that ignorance of fact is allowed to persist. The research and practice communities are like the proverbial Two Cultures[1]: separate, coexisting, wary of each other, and yet interdependent. At times the practice-research relationship seems to be like that between a monarch and his courtesan: for the monarch the courtesan decorates the court, is used to satisfy a need, and is ignorable in the affairs of the state. At the other end, to keep her position, the courtesan must cater to the monarch's needs and cultivate his good will.

In what follows I will describe instances of dysfunction in the practice-research relationship, state issues of concern, and point to the need for reform. Some will take exception to this jaundiced description, disagree with my observations, and differ with my conclusions. They will argue that over that last few decades good progress has been made towards evidence-based road safety management, and that this would not have been possible without the implementation of findings produced by dedicated researchers. If so, why then dwell on deficiencies and call for reforming an already satisfactory system?

[1] 'The Two Cultures' is the first part of an influential 1959 Rede Lecture by British scientist and novelist C. P. Snow. Its thesis was that "the intellectual life of the whole of western society" was split into the titular two cultures — namely the sciences and the humanities — and that this was a major hindrance to solving the world's problems.

My reasoning is simple. It is the 'State'[2] that produces and operates the road network and road-user safety is affected by how this is done. This is why those who on behalf of the 'State' plan, design and operate the road network should take road-user safety into account. However, in my view, their consideration of safety is too often based on opinion rather than on evidence. Inasmuch opinion is a fallible guide when it comes to the prediction of safety consequences, reliance on it is makes for unnecessary injuries, loss of life and waste of resources. My purpose is neither to praise nor to blame; it is to argue that the interest of road-user safety demands that, as much as is practicable, the prediction of safety consequences should be based on evidence, not on opinion. Because it is currently not so, a reform of the research-practice relationship is necessary.

The paper revolves around two historical anecdotes. The first anecdote is about research findings that should have raised questions. But questions were not raised and opinion trumped evidence. The second anecdote is about how a national standard is being created by bending research results so that they seem to support a pre-existing policy. Instead of evidence-based policy we will get policy-based evidence[3]. Telling these stories allows me to raise several issues of concern.

Anecdotes are good for illustration and can stimulate debate. But to argue that a reform of the research-practice relationship is necessary one has to show that the dysfunctions are evident not only in the anecdotes but are widely prevalent. How to convince that what the anecdotes illustrate is not the exception, that similar dysfunctions are actually quite common?

I could add story to story and pile example atop example[4]. But this would only make an overlong paper even longer and still fall short of convincing. In the final account only informed readers can say whether, in the light of what they know, the issues I raise are of concern. Not everybody is an 'informed reader'. To explain, my optometrist tells me that it would be unsafe for me to drive if my visual acuity was less than 20/50. He is a trained and experienced professional. However, there is nothing in his training or experience to tell him whether this (Ontario) standard is based on some inherited but untested opinion, on the expediency of blame-the-driver populism, or perhaps on data-based evidence. This is why he is not an 'informed reader'. He would be one if he knew what research about this matter shows[5]. To weigh in on the question of dysfunction prevalence I count on the community of informed readers, those who know what the practice is and who know what research shows.

2 THE FIRST ANECDOTE: TWO STUDIES ABOUT THE SAFETY EFFECT OF EDGELINING

In 1957 Ohio initiated a program of painting edgelines on all rural two-lane highways that were at least 20 feet wide. At that time "*No prior research on pavement edge marking was available in Ohio*" (Musick, 1962, page 1). After the edgelining program was underway a randomised controlled experiment (an RCE) was devised to estimate its safety effect. Nine pairs of road segments (116

[2] By the inclusive term 'State' I mean the federal, state, provincial and municipal governments.

[3] A turn of phrase attributed to Anne Glover, the Chief Science Adviser to the President of the European Commission.

[4] See e.g., Hauer E (2000 a and b, 2007, 2016)

[5] Research shows (see, e.g., Owsley and McGwin Jr., 2010) that if there is a link between visual acuity and crash risk it is so week that, so far, it could not be reliably measured. If my optometrist knew this to be the case he might be sympathetic to my argument that there is a problem, that practice and research findings should be more closely aligned.

miles) were selected so that the roads in each pair were similar. One segment of each pair was selected at random and got edgelines; the other segment was left as is and served as 'control'.

In Kansas edgelining also started in 1957. As Basile (1962) attests, at that time several states already made wide-spread use of edgelines "based on an anticipated reduction in accidents and fatalities", and that "No research had been attempted in Kansas prior to this time to determine the effectiveness of this device or of its economic justification." (p. 80). Limited before-after comparisons done in 1957 and 1958 and looked promising. However, since these roads had "comparatively high initial accident experience rates" the observed accident reductions could have been the artifact of regression to the mean. To get more solid results, in 1959, an RCE (Randomized Controlled Experiment) was initiated. Here 29 pairs of adjacent road segments (^{3}84 miles) were formed and the decision which segment of the pair is to get edgelining was also taken at random.[6]

I will provide more detail about these two studies later. However, even this skeletal description already raises some questions. Why study the safety effect of edgelining after the decision to edgeline has been made and the program was running; should one not do the research first? Is it OK to embark on such a program on the basis of opinion, a belief that it will be good for safety?

2.1 Why study the safety effect?

From a distance of sixty years it is difficult to know what considerations motivated the conduct of these two RCE's. In Kansas, in Ohio, and in many other states, edgelining programs were well underway before the two RCEs commenced. Since the RCE findings were not used to decide whether to embark on a statewide edgelining program what was their purpose[7]? Musick does not explain[8]. Basile says that, in Kansas, the purpose was to confirm in a scientifically solid way what the preliminary and limited studies indicated; namely, that edgelines reduce accidents. That is, to demonstrate that the decision to edgeline all roads which was originally made on opinion-based anticipation was in fact a good one.

There is merit in this type of post-implementation evaluative research if it can affect practice, if the findings can help to decide whether to stop, modify or continue implementation[9]. But in this case it might have been difficult to say: "Stop edgelining, we may have made a mistake". It is therefore not clear why the Ohio and Kansas RCEs were undertaken.

I am told that many (most?) operating agencies consider the post-implementation evaluation of the effect of their programs to be a hallmark of sound management. It is therefore useful to distinguish between two types of evaluative research. One is the *'Policy and Profession-Serving'* evaluative

[6] I know of no other RCE (Randomized Controlled Experiment) the purpose of which was to estimate the safety effect of edgelining. In fact except for Stock et al. (1983), Fosser (1992) and Kallberg (1993) I do not know of any other RCE for any treatment, intervention, design standard or operational warrant that affects road safety. Why is that? Is there a substantive difference between road safety and medical research to make it a rarity in one circumstance and the gold standard in the other? I will leave the question unanswered because I gave my opinion about this question in "Enhancing Future CMF Research" downloadable from ResearchGate. (Kallberg studied the effect of post mounted delineators on nighttime speed and accidents and found that they increased both.)

[7] To gain insight I tried to contact persons then active in order find out what were the historical circumstances surrounding the conduct of these experiments. Unfortunately too much time passed and my efforts did not bear fruit.

[8] He only says that since the effect was not studied before the decision to edgeline all highways it was studied afterwards.

[9] The events described in Stock et al. (1983) are one such example. Since in this RCE the high school driver education programs was found ineffective the findings led to a decline in the popularity of such programs. The Fosser (1992) RCE is an example of the opposite. Even though in Norway periodic motor vehicle inspection was found to be ineffective it remained a requirement for entry into the EU.

research the purpose of which is to estimate the safety effect of a program or intervention in order to inform future actions, often elsewhere and by others. The other type is the *'Operating-Agency-Serving'* evaluative research the purpose of which is to inform the management of the agency about whether its program or intervention has achieved its goal and, if not, whether that program needs to be modified or discontinued. The 'Policy and Profession-Serving' evaluative research is the source of factual knowledge for the evidence-based road safety delivery. The 'Operating-Agency-Serving' evaluative research is a tool of management.

2.1.1 Issue 1: Two purposes of evaluation research

The difference of purpose between the two types of evaluative research should be kept in mind because, as I will show, the 'Operating-Agency-Serving' evaluative research can be subject to external considerations that may influence what is researched and color the findings. It may benefit an operating agency to engage in evaluative research if the discontinuation or the modification of an implemented program is an option. When this is not the case there is no good reason for an operating agency to engage in such research. It is not the usual role of an operating agencies to undertake 'Policy and Profession-Serving' evaluative research.

2.2 Between opinion and evidence

In the 1950s, when Ohio, Kansas and other states started edgelining their roads, it was not known how this will affect safety. At that time, before the vagaries of road-user adaptation became better known, one could easily surmise that that making the edge of the lane more visible is likely to help drivers to stay on the road and is not likely to do harm. In any case, 'opinion' was deemed sufficient to begin wholesale edgelining. Is this as should be?

When making a decision is urgent and factual evidence is missing opinion must suffice. However, in the edgelining anecdote there was no such urgency; it would have been possible the do the RCE's beforehand. Should we be concerned about a frame of mind in which opinion about what the safety consequences might be is sufficient?

The long-ago events of the anecdote would be of little interest were the same frame of mind not still with us; however, in some quarters it is. Consider, e.g., the Policy on Geometric Design and the Manual on Uniform Traffic Control Devices[10]. They are the product of committees and reflect the opinions of their members. One cannot take comfort in thinking that evidence-based consideration of safety is already appropriately incorporated in these standards and warrants (see Hauer 2000 a, b and 2016). Nor do these documents tell the highway designers and the traffic engineers how their choices affect safety. And so, professional decisions based on these documents reflect an agglomeration of opinions and interests without their evidence-based safety consequences being declared and, in some cases, without being known. And yet these are the decisions that determine the safety of our road system.

2.2.1 Issue 2: The road is a product

Times have changed. Nowadays the safety of products such as food, drugs, vehicles, appliances, etc. is not left to be determined by the opinion-based anticipation of their producers. Such products are usually allowed into use only after evidence of their safety is ascertained and provided[11]. A road is also a product. The magnitude of the danger inherent in road use is built into it by the design and operational decisions made by its producers. This is why, today, In keeping with what we expect of

[10] The last editions of which are AASHTO (2011) and FHWA (2009).

[11] Before a drug is allowed into public use one must have data-based research about its expected benefit and harm. Even as hundreds were dying daily of Ebola, vaccination was withheld till its beneficial and harmful effects could be ascertained.

other products, road-users may want to insist that the safety of roads also be based on evidence, not on opinion-based anticipation[12]. If reliance on opinion-based anticipation of safety consequences is to become less legitimate there has to be a change in the culture of agencies responsible for the production and operation of roads and in the professions that are their agents[13].

2.3 Opinion is a fallible guide

As already noted, one can hardly fault the decision-makers in Ohio and Kansas in the 1950s for thinking ('anticipating') that edgelining will improve safety. It was and still is common to believe that making the edge of the pavement more visible will prevent accidents. Did the RCE's confirm the 'anticipation' of the decision-makers in Ohio and Kansas that edgelining will help safety?

To answer I return to the Musick and Basile papers. Both found that following edgelining the number of accidents at access points (intersections, alleys and driveways) was much less than what could have been expected without edgelining. This, in itself, was unexpected because the mechanism to bring about such an effect was unclear[14]. Even more surprisingly, and contrary to what was expected, both studies found that between access points the number of accidents after edgelining was larger than what would have been expected without it. In Ohio"*…the net increase in* (these) *accidents was approximately 15%...*" (Musick, page 4); in Kansas "*… a 27% net increase in* (these) *accidents … is found.[15]*" (Basile, page 83).

One might defend the reliance on opinion-based anticipation if it were nearly always correct. Unfortunately, in road safety, belief, judgment, and opinion-based anticipation are fallible guides. Thus, e.g., most did not anticipate that driver education in high schools will prove unhelpful to safety, that periodic motor vehicle inspections do not save accidents, or that placing raised pavement markers on curves will increase the chance of crashes. For many other failures of anticipation see Smiley (2008) and Rudin-Brown & Jamson (2013). It is now clear that one may not trust simple principles such "better visibility makes for fewer accidents"[16]. Human behavior is complex, adaptation is one of its hallmarks, and the safety consequences of interventions are difficult to anticipate[17]. That in the Ohio and Kansas RCEs accidents decreased at access points

---

[12] One could argue that it is feasible to determine the safety of food, drugs etc. before they are allowed into use whereas the same is not possible for roads. While the difference is real, one can easily think of a safety testing protocol for roads. Before some road design or treatment is put into use it could be tested on simulators, investigated in pilot projects, tested in larger scale randomized controlled trials etc.

[13] See e.g. Hauer, E, (2007).

[14]There are no edgelines at access points since there is no edge of pavement to be made visible. Basile speculated that with edgelining away from access points drivers might be looking further ahead or that the termination of the edgeline before the access point might make its presence more obvious. He suggested that "*Carefully planned research is needed to test … these theories.*" (Page 83).

[15] Basile adds that*:* "*This net change (the 27% increase) has a low level of significance, however, 0.17. In other words the net increase of 27 percent could easily have occurred by chance alone.*" (p. 83). As written, the statement is incomplete and likely to mislead. A more complete and correct wording would have been: "*If edgelining did not change the probability of accident occurrence then there would be a 0.17 probability to observe a 27% or larger increase in the count of accidents.*" Of course, a 17% chance of something occurring implies an 83% chance of it not occurring. Therefore the phrase "could have easily occurred" must be taken with a grain of salt. The purpose of Basile's statement is to blunt edge of the finding that edgelining between access points is likely to do harm.

[16] It is therefore surprising and disappointing that, as I will relate in the second anecdote, the FHWA predicated its case for coining a minimum retroreflectivity standard on the same simplistic and fallible premise - that improving edgeline visibility reduces accidents.

[17] To illustrate the complexity and difficulty consider the findings of Kapoor and Magesan (2014) who write about the safety effect of installing pedestrian countdown signals in Toronto. The signals are intended to assist pedestrians, to modify their behavior, and thereby to reduce pedestrian injuries. However, apparently, there was also an unintended consequence, drivers

and increased in-between speaks volumes. It is the exact opposite of what one would opine and anticipate.

2.3.1 Issue 3: In road safety opinion-based anticipation is insufficient

No amount of personal experience enables one to reliably predict the safety consequences of safety-related decisions; only data-based research can do so. In unique situations the exercise of judgment in decision-making is inevitable. But when a decision applies in many circumstances (as e.g. the edgelining all roads or the periodic inspection of motor vehicles) reliance on opinion carries a large risk of harm to road users and/or a waste of resources.

2.4 Road users have rights

Truth can be inconvenient. Research findings can produce results that cast doubt on the wisdom of previously made decisions. This is what happened in Ohio and Kansas. Had it been suspected beforehand that painting edgelines between access points may do harm it would have been prudent to do more research before full-scale implementation. However, the results of the RCEs came to light only after the edgelining programs were already running for some years. What were the Ohio and Kansas DOTs to do at that point in time? To admit to a mistake in private life attests to strength of character; to admit to a mistake in public life is unpalatable, might be embarrassing, damaging to careers, and it could generate liability. And yet to just disregard what this research found was clearly wrong.

I tried to find someone to ask about these events that unfolded more than half a century ago but did not succeed. Therefore I cannot know whether the higher-ups knew and understood what their in-house research showed. Still I have to point to the elephant in the room. Two high quality studies independently found that, contrary to what was expected, edgelining between access points seemed to do harm. This did not stop the edgelining programs then in progress. By disregarding what the in-house research found some Ohio and Kansas road users may have been harmed[18].

---

too changed their behavior. Thus, the authors claim that: "… we find that (vehicle-vehicle) collisions rose largely because of an increase in tailgating among drivers, a finding that implies drivers who know exactly when traffic lights will change behave more aggressively."

[18] Commenting on an earlier draft of this paper some Australian colleagues pointed out that that what may have been common in the 1950s may not a big problem now; that nowadays evidence is less likely to trump opinion and road user safety is less likely disregarded. This is not what I see in North America. I think that while progress has been made the problem persists. While contemporary examples of such disregard abound it is (again) not practical to demonstrate prevalence by listing them. The best I can do is to add a contemporary story to what already has been said in issue 2. (There I claimed that roads can be made safer or less safe and that by following road design standards and traffic engineering warrants the level of safety provided to road users is unpremeditated). A determination of how pervasive the neglect of the road users' right to safety still is must emerge, as before, from the collective wisdom of 'informed readers'.

The contemporary example is about the installation of raised pavement markers. Several local studies of their effect on safety resulted in contradictory estimates. Yi (2006) lists seven research studies with estimates ranging from 26% decrease to a 30% increase in nighttime crashes .To come to a consensus opinion the National Cooperative Highway Research Program sponsored a large research project that used sound methodology on plentiful observational data from six states. The study (Bahar et al., 2004) found, e.g., that contrary to common belief and practice, the installation of raised pavement markers on two-lane rural roads is associated with an increase in crashes, especially on sharp curves. The estimates from this project were eventually included in the Highway Safety Manual (AASHTO, 2010). Yi (2006) contacted several State DOT "*to see if they have adopted or plan to adopt the NCHRP guideline for raised pavement markers* (those in Bahar et al. 2004). He reports that *"Iowa, Pennsylvania*[18]*, and Michigan indicated that they do not plan to adopt the NCHRP guidelines."* (pp. 37-38). As in the edgelining anecdote, disregard of what research found may be harming road users.

If a similar situation arose in the pharmaceutical or in the food industry (not to mention in the commonly reviled tobacco industry), if it turned out that inconvenient research evidence was disregarded, the public outcry would be loud and the retribution harsh. Why is it different when the product - the road - is designed, built, and operated by the State?

One part of the answer[19] is a culture that values action, trusts opinion, and cherishes the decision-making freedom that the discounting of evidence allows. As discussed in Section 2.2 that culture is still with us. Additional support for this claim will emerge in the second anecdote.

If the road users knew that in the planning, design and operation of the road network by the "State" their safety is still allowed to be determined by opinion rather than evidence, they might object. But the road users do not know that. They are led to believe that their own transgressions are the cause of accidents and therefore, by faulty logic, that prevention is necessarily by training, education, legislation and enforcement. This archaic notion and logical misconstruction[20] is perpetuated by the police, the media, the tort law and by, in North America, the 'State'.

2.4.1 Issue 4: The right of road users

Roads can be designed and operated to be safer or less safe. The responsible 'State' should strive to make safety-related decisions on the basis of fact, not opinion. To this end it may help to formally declare that road users have a right to safety (see, e.g., Mohan, 2003). Specifically, that:

Road users have a right to expect that 'State' decisions that substantially affect the safety of many should take into account the fact-based expectation of the safety consequences of such decisions[21].

I am not competent to say how to change the prevailing culture so as to adequately protect road user rights. But the matter should be aired[22]. For, if not, if research is to be subservient to considerations of publicity and liability its trustworthiness and usefulness will be limited, its role ambiguous, road users will be unnecessarily harmed and their right to safety will continue to be abridged.

2.5 The shackles of dependence

I now return to the story of the two RCEs. Because theirs was an 'Operating-Agency-Serving' kind of research[23], Musick and Basile worded their conclusions circumspectly. The benefit of edgelining was highlighted and its harm hidden from sight. How was this done?

In summarizing the principal findings (on page 5) Musick speaks only about those circumstances in which edgelining was associated with a statistically significant decrease in accidents and then

[19] Another part is the doctrine of Sovereign Immunity (see, e.g., Chemerinsky, 2001).

[20] Even if the cause of cholera is a bacterium its prevention in the 1854 Broad Street outbreak in London was by removing the handle of a street pump. Similarly, even if the cause of some run-off-the-road accidents may be driver fatigue, their prevention may be by rumple strips of guardrails.

[21] O'Neill (2002) argues convincingly that rights and obligations are two sides of the same coin and prefers to anchor human rights in human obligation. Here the obligation is of the 'State' and its employees to do the right thing by road users by making safety-related decisions on the basis of fact, not opinion. Just as individual autonomy is the cornerstone of bioethics, so the right of the road user to evidence-based consideration of safety consequences is the foundation of ethical considerations in transportation.

[22] One of the key elements of the Vision Zero and Safe Systems doctrines (see, e.g., ITF, 2016) is that not only the road user but also those who design, build and operate roads are also responsible for preventing accident occurrence.

[23] See 'Issue 1'.

proceeds (on pages 6 and 7) to list his four conclusions. All four are about specific accident reductions. That between access points there was a 15% increase in accidents goes unmentioned. It was swept under the carpet in the second conclusion by saying that: *"Accidents at intersections, alleys, and driveways were significantly reduced but accidents between access points showed no significant change"*. Because in English the synonym of 'not significant' is 'unimportant' the non-statistician reader is easily misled to think that in Ohio edgelining was found to be an entirely unmixed blessing; that it only reduced accidents.

When Musick said "*…no significant change*" he meant[24] (but failed to say) that the 15% increase was not 'statistically significant'. But statistical significance has nothing to do with significance in the ordinary sense. I do not want to be sidetracked into the labyrinthine reasoning required to explain what statistical significance does mean. Suffice to say that in an authoritative statement the American Statistical Association (Wasserstein and Lazar, 2016) came out strongly against these kinds of misinterpretations saying that*:* "*… A p-value, or statistical significance, does not measure the size of an effect or the importance of a result. Statistical significance is not equivalent to scientific, human, or economic significance. Smaller p-values do not necessarily imply the presence of larger or more important effects, and larger p-values do not imply a lack of importance or even lack of effect.*" (p. 132). The substantive error is in thinking that absence of statistical significance says something about the presence or absence of change in the probability of accident occurrence; it does not.

Simply said, the most likely inference from the Ohio data was (and is) that after edgelining between access points accidents increased by 15%. But saying so without obfuscation in a published report might strain collegial relationships, test organizational discipline, and court the danger of censure. And so Musick used the "*…no significant change*" phrase to instead of saying what the data really showed.

In Basile's paper about the Kansas RCE the conclusions are also four. All four, again, speak about accident reductions. Only the third conclusion mentions accidents between access points and reads: "*Accidents at intersections and driveways were significantly reduced during both daytime and nighttime conditions. Accidents between access points were not significantly changed.*" (Page 86). Here too the reader is left with the (false) impression that edgelining is all benefit and no harm. But when one reads the body of the paper the opposite appears to be closer to truth. About three quarters of all accidents occur away from access points and these, after edgelining, are said to have increased by 27% (Basile, 1962, Table 6, page 85). Total accidents increased by 1% and the number of persons killed and injured increased by 16%.

In both papers the verbal device for obscuring this truth was that of not distinguishing between significance in the ordinary (i.e. importance) and statistical significance. In both papers the discrepancy between the message of the data and the words written into conclusions is clear. Given the circumstance in which these operating-agency-serving evaluations were done, the discrepancy is understandable. But its net effect is that of misinformation, of stripping data of their message.

I distinguished earlier between two types of evaluative research and cautioned that 'Operating-Agency-Serving' type can be subject to various external considerations that may color its findings. This is an example. The story told in the second anecdote will further exemplify noxious connection between the interest of the operating agency and the research it undertakes or finances.

[24] See footnote 2 to his Table 3.

2.5.1 Issue 5: The importance of research independence

I am led to observe that for research to be the engine of progress towards evidence-based road safety management it must be effectively insulated from the pressures of past decisions and present practices. For, if not, some findings will tend to be worded to suit interests of sponsoring organizations and decent people will be placed in untenable positions. Placing the research function within agencies that abhor bad publicity and dread exposure to liability will motivate the desire to exercise control over what is researched and what is openly reported. Making of such agencies into custodians of research budgets gives them complete control. Such control is the nemesis of useful research and an impediment to learning from experience.[25]

2.6 An interim summary

I set out to describe problems of coexistence between research and practice in North America. The edgelining anecdote based on the Musick and Basile papers suited my purpose admirably. Randomized controlled experiments are the gold standard for research of this sort. Therefore the findings could not be dismissed as methodologically faulty. In addition, the findings of one were independently replicated by the other. Most importantly, all my observations and claims could be substantiated from within the text of published papers. I used the anecdote to highlight some signs of dysfunction in the relationship between research and practice and to raise the related issues.

The weakness of this anecdote is that it is old. Its bite could perhaps be diminished by saying that the dysfunctions present more than half a century ago are not typical of today. This objection cannot be sustained for long. The next anecdote is entirely contemporary and, as I will show, it exhibits the same dysfunctions, only on a larger and more worrisome scale. My own life in road safety spans most of the period between both anecdotes. During this time I have been rubbing shoulders with decision-makers and researches. From what I have seen, plus ça change, plus c'est la même chose. Decisions, policies, standards, and warrants that affect life and limb are still influenced by opinion-based anticipation as well as public relations and liability considerations and these still play a role in deciding what is and what is not researched, what is being reported and what is withheld. To the extent that my perception is correct thought should be given the listed issues, and perhaps to some that I failed to list. By confronting the strategic question of how to better structure the relationship between research and practice both can be improved.

3 THE SECOND ANECDOTE: RULEMAKING ABOUT PAVEMENT MARKING RETROREFLECTIVITY

In this anecdote the issues are broader and writ on a larger canvas. While the first anecdote played out in the distant past this one is unfolding as I write. Here is the story.

I stumbled on a 'Notice of Proposed Amendment' (NPA) in the Federal Register[26] (2010) while reviewing what is known about the link between pavement marking retroreflectivity and safety

[25] Lackey (2016) would call research produced in such setting 'normative science' which he defines as "information that is developed, presented, or interpreted based on an assumed, usually unstated, preference for a particular policy choice." He calls the use of normative science "stealth advocacy" because those reading such statements are likely to be unaware of the hidden advocacy. He says that "it is a corruption of science and should not be tolerated in the scientific community – without exception".

[26] The official journal of the federal government of the United States that contains government agency rules, proposed rules, and public notices

(Hauer, 2017a). In this NPA the Federal Highway Administration (FHWA) responds to congressional legislation that directs the Secretary of Transportation to revise the Manual on Uniform Traffic Control Devices (MUTCD)[27]so that it requires some minimum levels of pavement marking retroreflectivity that must be maintained (Congress, 1992). The NPA contains dismissive comments about research to which I made a contribution (Bahar et al., 2006). Was our research bad and were its conclusions wrong? If not, why did the FHWA disparage good evidence? Besides, why did Congress spend legislative time on such an esoteric topic? Does Congress legislate, say, the minimum dose of Statins to treat elevated cholesterol? Are these not decisions that professionals should make? My vanity was hurt, my curiosity piqued, and I decided to investigate. In the course of this investigation which began for entirely unrelated reasons I learned much about how research is used to support policymaking.

My search for the origin of the Congressional interest in retroreflectivity led to the 'Hearings Before the Subcommittee on Surface Transportation' of the 102nd Congress (1991). Appearing before the subcommittee was the President of the American Traffic Safety Services Association (ATSSA). ATSSA is a trade association whose members supply most of the traffic control equipment and services (including pavement markings) on US highways[28]. In his submission (Statement, 1991), under the heading of 'Improved Safety for Aging Drivers'[29], the president of ATSSA recommended that: "*Congress should direct the Secretary to propose rulemaking … to require minimum maintained levels of retroreflectivity…*" (Page 1503). What ATSSA asked for the Congress made into law. To wit*: "The Secretary of Transportation shall revise the Manual of Uniform Traffic Control Devices to include … a standard for a minimum level of retroreflectivity that must be maintained …"* (Congress 1992). This was passed on to the FHWA for 'rulemaking' and thereby led to the aforementioned NPA[30].

One may be uneasy about Congress passing legislation that was proffered by a trade association, ostensibly out of concern for the safety of older drivers. However, such a sentiment is not apt as it is the task and prerogative of legislators to assess the merit of a proposed legislation, whatever its source, and to do so on the basis of their best judgment. The judgment of legislators does not require quantified evidence about cost and benefit[31]. However, when legislation is to be translated into actionable 'rules' then, by Executive Order 12866, the executive branch is obliged to consider potential costs and benefits[32]. Here the intended benefit was improved safety and therefore the task of the FHWA was to determine what the safety benefits of maintaining some minimum level of pavement marking retroreflectivity are[33].

---

[27] The MUTCD is a document issued by the Federal Highway Administration of the United States Department of Transportation specifying the standards by which traffic signs, road surface markings, and signals are designed, installed, and used.

[28] According to Carlson et al. (2009) in 2007 the nationwide annual pavement marking expenditure was about $2 billion, not small change.

[29] Even though on the written page ATSSA seems concerned about the safety of older drivers, it must have been obvious to the members of the Subcommittee that what is at stake is the sale of paint and glass beads.

[30] 'Rule' or 'regulation' making *is "one of the basic tools of government used to implement public policy.*" (Carey, 2013, Summary)

[31] I am told (on high authority) that the benefit and cost of regulatory legislation is rarely estimated (by anyone) prior to passage.

[32] Carey (2014) summarizes what analysis is required in rulemaking. Chief among Carey's references is the Executive Order 12866 (Federal Register, 1993). That Order says that Cabinet Departments "*should assess all costs and benefits of available regulatory alternatives, including the alternative of not regulating.*" (Section 1.)

[33] And here is the rub. I am told (also on high authority) that the circumstance in which the executive branch cannot find sufficient benefits for what Congress legislates arises with some frequency. What then does the executive

At first glance the task may have seemed unproblematic. It stands to reason that improving the visibility of pavement markings cannot but help the driver to stay on the road. However, as already noted earlier, road-users adapt to circumstances and this may complicate matters. Drivers may adapt visibly by adjusting their speed and they may adapt invisibly by, say, being less alert. How people adapt is not easy to anticipate; they may over-compensate, under-compensate, or keep the risk constant. As already noted, simplistic principles have proven to be fallible guides for predicting the consequent change in accidents. This is perhaps why it proved difficult to find research showing that more pavement-marking retroreflectivity makes for safer roads. Even today, as Carlson et al. (2015, Abstract) write, "*The relationship between pavement marking retroreflectivity and nighttime safety has been a topic of research for past decade or more but consistent findings have been elusive… .*" The same inconsistency of findings characterized the state of knowledge in 1992 when the legislation landed in the Transportation Secretary's lap.

3.1 The 'Synthesis' in the service of rulemaking

What to do when the task is to provide evidence supporting a pre-existing decision but the evidence is murky[34]? One way out is to create a Potemkin Village by exaggerating the credibility of favorable research findings while detracting from the believability of inconvenient ones. To this effect, as I will show, "*In 2008, the FHWA developed a synthesis of the benefits of pavement markings, including safety studies."* (Federal Register, 2010, page 20936.) I will refer to this document as the 'Synthesis' (see Carlson et al., 2009).

The 'Synthesis' was to review what research says about the relationship between pavement marking retroreflectivity and safety and thereby to show that the extant evidence supports the proposed rule. However, perhaps because direct evidence was sparse and inconclusive, the 'Synthesis' sought support for the rule in related matter.

One such related matter was edgelining. If edgelining could be shown to be good for safety, so the argument might go, then making sure that edgelines stay visible must also benefit safety. To this effect the 'Synthesis' leads its readers to believe that *"robust"* research found edgelining to be good for safety, and that retroreflectivity matters because edgelining saves only nighttime accidents. However, this is not what can be fairly concluded from the research reviewed in the 'Synthesis'.[35] Chapter and verse about the discrepancies between what the 'Synthesis' says about the safety effect of edgelining and what the reviewed research evidence actually shows is in Appendix A. Here I provide only a summary.

The study which the 'Synthesis' called "*one of the most robust"* is in fact based on bad data and has at its core a false assumption. I do not think that even its authors would call the findings robust. Another study which the 'Synthesis' calls *"most often cited"* concocted a fanciful estimate of the

---

branch do when it is difficult to show that enacted legislation is likely to be beneficial? It could ask Congress to reconsider. But doing so is unpalatable and, so I am told, would be fruitless. The more practical option is to bend with the legislative wind and unearth sufficient benefits by creative use of the extant research evidence.

[34] At this point it is tempting to raise the 'large canvas' issue of what is missing from the 'checks and balances' put into the constitution in the 18th century when it comes to governing in the 21st century. Can the executive branch that is ordered to quantify costs and benefits go about its rulemaking job honestly when the legislators use only un-quantified judgment? I will not go down this road in order to keep to the much narrower issue of honesty and usefulness in road safety research.

[35]The economist John Kenneth Galbraith coined the term 'Innocent Fraud'. He used it to describe a lie or a half-truth that, because it suits the needs or views of those in power, is presented as fact. After much repetition, the fiction becomes common wisdom. *'It is innocent because most who employ it are without conscious guilt,'* Galbraith wrote in 1999. '*It is fraud because it is quietly in the service of special interest.'*
http://www.thirdworldtraveler.com/Economics/FreeMarketFraudGalbraith.html

safety effect of edgelining from the findings of another study disregarding the cautions of that study's authors. From two respectable research studies the 'Synthesis' culled only those results which suited its mission and kept mum about findings that negated it[36]. In sum, the upbeat message of the 'Synthesis' about the safety benefits of edgelining is made up of creative interpretations of the studies it chose to review[37].

Having written the marching orders for the 'Synthesis', the FHWA used these 'creative interpretations' to justify the problematic rulemaking. It says in the Federal Register that: *"The safety benefits of adding edgelines was demonstrated for nighttime conditions, low visibility conditions, and highways with narrow pavement widths and low traffic volumes."* (Page 20936. Emphasis added.) This is untrue. The 'Synthesis' does not demonstrate what the FHWA claims it does.

Another 'related matter' reviewed in the "Synthesis' is about how edgelining affects speed. If drivers reacted to the addition of edgelines by modifying their speed it would be difficult to justify the no-effect-on-daytime-accidents assumption on which the findings of the *"most robust"* study rest. More importantly if edgelining induced an increase in speed it could harm safety. To forestall these twin dangers to the rulemaking, the 'Synthesis' reviews in a separate section (pages 10 and 11) the findings of three research reports. About these the 'Synthesis' claims that *"…most show no significant effect in absolute speed difference or, perhaps more importantly, speed variance."* Actually the opposite is true. As shown In Appendix B all three studies show that edgelining does affect speed, that speed is affected both during the day and night, and that the magnitude of the change depends on circumstances (straight road or curve, amount of traffic, width of lane etc.) Here again the 'Synthesis' constructs a virtual reality that is at odds with what the reviewed research shows.

For the FHWA and its rulemaking duty the most irksome research finding must have been that by Bahar et al. (2006). The basic idea behind that study was simple. The retroreflectivity of pavement markings fades over time and is periodically restored by repainting. One can therefore check how the accident frequency of roads changes as a function of the decline in retroreflectivity keeping all else constant. Were one to find that accident frequency diminishes just after repainting and then increases as retroreflectivity declines that would be a sign that retroreflectivity matters. But Bahar at al. found that road safety does not improve just after the markings are repainted nor does it decline as retroreflectivity degrades over time. Such a finding is an anathema to what Congress instructed the FHWA to do. The FHWA must have hoped that the 'Synthesis' will find something wrong with the Bahar et al. findings The authors of the 'Synthesis' obliged by saying that the Bahar et al. research has *"significant limitations*" (on page 12) and that "*Combined, these limitations and concerns seriously challenge the quoted concluding remarks …."* (On page 13)[38].

As I show in Appendix C four of the five alleged limitations have no merit. The fifth limitation which the 'Synthesis' notes is that Bahar et al. (2006) used data from California where retroreflectivity is only seldom less than 100 mcd/m^2/lx and therefore the no-effect-on-safety finding applies only to

[36] Thus, e.g., they do not say that Musick and Basile found edgelining beneficial at access points and harmful in-between, that edgelining in Kansas was associated with an increase in nighttime accidents, etc.

[37] I do not claim that edgelining does not benefit safety because I read only those reports that are referenced in the "Synthesis'. Besides, if not for safety, edgelining may be justified on other grounds.

[38] Carlson et al. (2013, page 59) repeat these criticisms and amplify by saying that they are sufficiently grave to *"limit (the) acceptability"* of the zero-effect conclusion. Because I contributed the method by which Bahar et al. produced their findings, the criticism in the 'Synthesis', in the Federal register, and its repetition by Carlson got my goat.

higher retroreflectivity values. This is a fair observation at least for white pavement markings.[39] Of course, the obverse is also true. Since data with white marking and retroreflectivity above about 100 mcd/m^2/lx and with yellow markings above 80 mcd/m^2/lx were plentiful, in this range one can say with some confidence that a decline in retroreflectivity was not associated with an increase in accident frequency.

It is of course possible that when retroreflectivity is very low there might be an increase in accident frequency but such a speculation goes against the Synthesis' own reasoning[40]. As is well known, the larger the retroreflectivity the better are visibility measures such as 'detection distance', 'preview time', 'visibility rating' etc. The 'Synthesis' asserts (on page 12) that *"Detection distances are thought to be a surrogate for crash data in that longer detection distances have a positive effect on vehicle-control measures and, consequently, crashes."* If so, how is it that increasing retroreflectivity (and thereby detection distance) above the limiting value does not benefit safety (as Bahar et al. have demonstrated) but decreasing it below that limit, as the speculation suggests, increases accident frequency? Usually 'Natura non facit saltum'[41].

Based on the criticism written into the Synthesis the Federal Register says that the Bahar et al. study *"… provides little if any information regarding the link between minimum pavement marking retroreflectivity and safety."(*Page 20936). This, as far as it goes, is true. But, taken in context, it is a peculiar statement. The business of the register is to show that allowing retroreflectivity to slip below some minimum is bad for safety. Bahar et al show that there is no evidence for such a decline all the way from fresh (white) markings till about 100 mcd/m^2/lx. Is that not relevant information? Given the success of the method used by Bahar et al. and if the FHWA really wanted to determine whether low retroreflectivity harms safety, would it not have been sensible to look in more detail in the low retroreflectivity data of Bahar et al. (2006) or do a similar study with more low retroreflectivity data? Nothing of that sort was done. Given that above some minimum more retroreflectivity does not buy more safety and that below that minimum the association of the two is unknown, can one regard the proposed rule justified?

This paper is not about the safety effect of edgelining, nor about the role of retroreflectivity or the merit of the proposed rule; it is not even about the tension between the impression which the 'Synthesis' tried to create and reality. This paper is about how road safety research findings are use in practice. This is what the rulemaking anecdote illustrates. Important issues arise.

I will distill the lessons of the Second Anecdote into three sets of issues. First there seem to be obstacles to learning from experience. Why with more than half a century of edgelining and putting glass beads in pavement markings was their safety effect in 2009 still so poorly known? Second, the indication that in some circumstances edgelining harmed safety was allowed to lie fallow. What to do about the ease with which such research findings can be disregarded? Third, research findings are often produced and used to support of pre-existing interests or policies[42]. What are the consequences of the closed loop between funding and finding? All three sets of issues bear on the relationship between research and practice.

---

[39] For yellow markings the lowest retroreflectivity bin was 15- 79mcd/m^2/lx and here too Bahar at al. found no effect on accident frequency.

[40] The 'Synthesis' has a section entitled "Studies of Visibility in Terms of Detection Distances" (Pages 13-14).

[41] Nature does not make jumps.

[42] Anne Glover, the Chief Science Adviser to the President of the European Commission lamented that evidence-based policy too often turned into its opposite, policy-based evidence.

3.2 First set of issues: Obstacles to learning from experience

The 'Synthesis' claims that there is evidence-based consensus about the beneficial safety effect of edgelining. But when I scratched the surface it turned out that in 2009 (when the 'Synthesis' was published) the referenced evidence was old, based on poor data, and (with the exception of Musick and Basile's work) obtained by deficient methods. If there was consensus its foundations were shaky. After half a century of edgelining the extent to which it affects safety in various circumstances remained poorly known. This is deeply disconcerting. Not learning from experience, not recognizing the need to do so, and not having a functioning system for the creation of defensible knowledge are aspects of a learning disability.

The learning from experience I speak of is a process that has, as shown in **Figure 1**, two distinct parts. Part I is that of 'Evidence Creation': the safety evaluations of programs, treatments, interventions or decisions that are undertaken by various researchers, at different locations and times. Part II is that of 'Representing Current Knowledge' in which, at some point in time, the information accumulated in the course of 'Evidence Creation' is examined and boiled down to something practically usable.

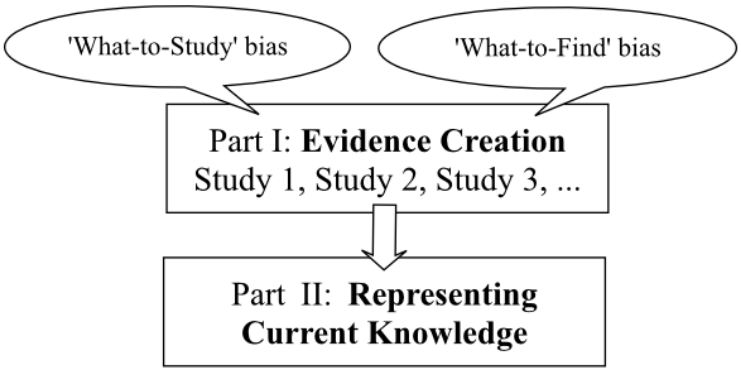

Figure 1. The learning from experience process

The dysfunctions in Part I come about when safety evaluations are initiated and funded by bodies that have an interest in the findings. This problem emanates from two junctures. The first juncture is when the question of what to study is contemplated. The downside of initiating an evaluation study is that it might show that something that was hoped for did not pan out. One may therefore expect that iffy programs and interventions are less likely to be evaluated than those where finding success is likely. Call it the 'What-to-Study' bias. The other juncture occurs when, perhaps because evaluating is mandated or in the expectation of success, an evaluation is done. These findings will be created and vetted with an eye to profit, publicity, liability and similar considerations. It is at this stage that convenient measures of performance can be chosen, inconvenient associations disregarded, favored conclusions given prominence, and ill-suited ones de-emphasized, explained away, concealed[43]or even shelved. This is the 'What-to-Find' bias. The choices made at both

[43] I describe an amusing concealment episode in Hauer (2000a, page 17). Researchers found that roads with 23' or wider pavement had more accidents than identical roads with 21'-22' wide pavements. This was contrary to the AASHTO Policy on Geometric Design of Rural Highways. So as to avoid the appearance of a conflict between what research found and what the standards mandate the researchers merged some pavement width categories and made the appearance of conflict vanish. Without digging deep into the report the trick would have been difficult to spot. In another episode a government agency evaluated the effect of the legislation which made daytime running lights mandatory. The analyst could choose amongst several comparison groups. He chose that which made the legislation look (just) cost-beneficial. Had a different comparison group been chosen it would not meet the benefit

junctures harm the learning-from-experience process by opening the door to bias. The bias in Part I flows into Part II.

Bias in estimates of the size of a safety effect harms road users. When the size of the effect is exaggerated, resources spent could be used to more effect elsewhere. When the size of the effect is under-estimated too little is spent to reduce harm. To eliminate the twin sources of bias in **Figure 1**, to protect the road user's right to safety, profession-serving evaluation research needs to be separated from agencies in charge of execution.

3.2.1 Issue 6: Separation of profession-serving evaluation research from execution

It is the FHWA, the state DOTs and the large municipalities that make the decisions and coin the standards and warrants by which roads are designed, built and operated. This is why these public bodies are justly concerned about what evaluation research may find and are therefore not a suitable setting for profession-serving evaluation research. To have appropriate checks and balances, to insure the impartiality of research, to guard against bias, the functions of profession-serving evaluation and of execution need to be separate. It is in this spirit that the NHTSA was set up not to depend on the automobile industry and the FDA to be separate from the pharmaceutical industry. I touched on the need to separate evaluation from execution in Issue 5. Failure to do so will continue to obstruct learning from experience.

Part II of the leaning-from-experience process, that of 'Representing Current Knowledge', is most often associated with the term 'meta-analysis'[44]. When the 'Evidence Creation Process' is tainted by bias so will be the meta-analyses based on it. One reason is that meta-analyses cannot see evidence from evaluations that were not undertaken. The direction of this 'What-to-Study' bias is easy to foresee; evaluations in which one could anticipate favorable findings will be over-represented in meta-analyses[45]. The other reason is due to evaluation studies that did see the light of day. When evaluation is not insulated from execution favorable findings will tend to be created and prominently placed in the abstract and the conclusion sections of reports (from where meta-analyses tend to extract them) while unfavorable findings will tend to be less conspicuously placed and therefore are less likely to find their way into meta-analyses.

3.2.2 Issue 7: Meta-analyses exaggerate safety benefits

When meta-analyses use findings that that are subject to the two aforementioned biases, they will tend to create belief in biased estimates of safety effect and thereby lead to inferior practice.

---

target. Here too the concealment was difficult to recognize. Later on the 'findings' of that study were used as a relevant fact when the same kind of legislations was considered in the EU.

[44]For the time being I will use the phrase 'meta-analysis' inclusively to signify also exercises such as 'reviews', 'syntheses', lists of crash modification factors etc. in which the accumulated evidence from several sources is assessed and an overall characterization of 'safety effect' is distilled.

[45] The 'What-to-Study' bias is distinct from the so called 'Publication Bias'. The latter arises after the research was done, when its outcome is known, and when making the outcome public is thought to be undesirable for some reason (no significant effect, contrary to what is 'known', etc.). The former arises before the outcome is known and there is anticipation that, were it known, its publication might be undesirable. The 'What-to-Study' bias pertains to the set of all implemented decisions and interventions while 'Publication Bias' pertains to a subset of these. Both biases affect meta-analysis adversely. For the 'Publication Bias' there exist partial remedies. It is not clear to what extent similar remedies could be used to diminish the harm of the 'What-to-Study' bias.

The 'Regression-to-Mean' and the 'What-to-Study' are both 'selection biases'. In the former the decision to treat is influenced by the outcome variable, in the latter the decision to evaluate is influenced by an anticipation of the outcome of the evaluation which may be correlated with the outcome were the evaluation done. Because it makes sense to treat sites with many accidents and because it makes sense to evaluate treatments that promise favorable findings both selection biases go in the same direction. Both tend to exaggerate the effect of treatments.

There is another dark side to meta-analyses; they separate practitioners from the original evidence and tend to create an illusion of truth by serial copying and repetition. And so, once an apparently stable estimate of safety effect is listed in a few meta-analyses, it may seem that all is well and that further research is unnecessary. Here is how this worked.

When saying that the pavement markings cause a 21% reduction in accidents, the 'Synthesis' (2009) copied this number from the meta-analysis in Miller (1991). Miller, in turn, came to the 21% estimate (somehow) using the conclusions in Basile, Musick, a table in Bali et al. (1978), and the list in Potters (or Macy et al., 1980). The Potters list, in turn, has been extracted (mostly) from original reports. It took me many weeks of pestering librarians and inter-library loans to assemble a near-complete set of the reports referenced in Potters. Few users of a meta-analysis will do the same. Fewer still would read the original reports even if available and only a small minority of those is in a position to assess the trustworthiness of the results. And so, by just seeing the 21% mentioned and referred to in several places[46] and with the sources practically unavailable, the reader of the 'Synthesis' can easily believe the 21% estimate is what the effect of edgelining is, that this number has gravitas and authority. As I show in Appendix A, the opposite is true[47].

Meta-analyses create a practically unbridgeable distance between the estimates they list and the evidence from which these come. Repeated exposure to such estimates creates a semblance of trustworthiness and, over time, the estimates acquire a ring of consensus. This is related to the 'Truth Illusion Effect' experimentally shown to exist by psychologists[48]. Practitioners have no choice but to rely on what the meta-analyses provide and act on this basis.

Once a meta-analysis is published practitioners have what they need. The passage of time, the chain-quoting of one meta-analysis by another, the repeated use of the listed estimate, and the unfamiliarity with the original evidential base, all tend to reinforce the belief that what is listed is trustworthy. If so, then doing more research will seem unnecessary. In this manner meta-analyses act as blinkers that shield practice from recognizing the flimsiness of its evidential foundation.

I asked why weak and occasionally disconcerting estimates of the efficacy of edgelining that have their origin in old and mostly deficient studies were deemed a sufficient basis for decades of unquestioning practice, and why the need for defensible evidence was not recognized. While the reasons are probably complex, meta-analyses may be one tile in this jigsaw puzzle; they can make it seem that what is needed is already available[49]. Given that the compilation of meta-analyses is a necessity, and considering its dark side, what is the remedy?

This is a difficult question to which I can give only a partial answer. It would be good if safety evaluation research could be cleansed of partiality and freed from the bias-generating dependence on agencies and their agendas. But since this is not likely to happen soon, a second-best option could be the making of meta-analyses a part of 'Systematic Reviews' as explained below.

---

[46] Miller (1991) is said to be *"one of the most often cited pavement marking safety studies."* (Synthesis, page 6),

[47] A similar chain of quoting occurred in the meta-analysis by Bahar et al. (2007) referenced in the 'Synthesis; it relied on the meta-analyses by Agent at al. (1996) and by Gan et al. (2005). I tried to identify the sources from which the Agent and Gan made up these estimates. One author did not remember what these were, the other did not respond to my inquiry.

[48] The illusion-of-truth effect is the human tendency to believe information to be correct after repeated exposure to it.

[49] Appending a standard error or a star rating to estimates in meta-analyses may help, but it does not remedy the two main problems. First, standard errors do not capture nor express bias and thus do not deal with the propensity of meta-analyses to exaggerate the safety benefits. Second, it is unclear how the psychological predisposition to trust what is repeatedly seen and often used is affected by a star rating or standard error

So far I lumped under the umbrella term 'meta-analysis 'exercises such as 'literature reviews', 'syntheses', lists of Crash Modification Factors etc. I now fold the umbrella and define meta-analysis properly as a statistical procedure for combining data from multiple studies. So defined, meta-analysis is but one of the activities within the broader concept of 'Systematic Reviews'[50]. Best known are the Cochrane Collaboration in health sciences and the Campbell Collaboration in the social sciences. To illustrate, under the auspices of the Cochrane Collaboration, Mulvaney et al. (2015) conducted a systematic review of the evidence on how changes in cycling infrastructure affect cyclist injuries. They found 21 studies looking at the effects of 11 different types of cycling infrastructure changes and concluded that "*Generally, there is a lack of high quality evidence to be able to draw firm conclusions as to the effect of cycling infrastructure on cycling collisions. There is a lack of rigorous evaluation of cycling infrastructure*".

Had the discipline of a systematic review been followed in the many meta-analyses, syntheses, reviews, Crash Modification Factor tables and the CMF Clearinghouse[51] that were published over the years, a similar conclusion would obtain about the safety effect of edgelining. A systematic review would find that, with the exception of Musick and Basile, all the other studies referenced in the 'Synthesis' are methodologically weak, that like in the cycling infrastructure quote *"There is a lack of rigorous evaluation …".* It would have been clear that the 21% must be questioned and is not an acceptable estimate for either rulemaking or practice. In addition, the need to embark on a research program that can provide acceptable quality evidence would have been recognized.

3.2.3 Issue 8: Systematic reviews, not just meta-analyses

By making use of evidence that is methodologically weak, based on poor data, and occasionally tainted by external interests meta-analyses tend to lead to inferior practice. The same deficiencies, when not identified, prevent us from recognizing the need for sound research. To make progress the example of the health and social sciences in the conduct of systematic reviews is worth following.

To this point I discussed two of the three aspects of the 'learning disability': not learning from experience and not recognizing the need to do so. Now I turn to the third aspect, that of the need for a planned and structured framework for the conduct of profession-serving evaluation research.

This lacuna is well exemplified by both the old and the contemporary anecdotes. As I show in Appendix A, several pieces of empirical evidence about the safety effect of edgelining emerged from a set of uncoordinated initiatives[52]. And so the various meta-analyses had to scavenge for

[50] Systematic reviews involve a detailed and comprehensive plan and search strategy derived a priori, with the goal of reducing bias by identifying, appraising, and synthesizing all relevant studies on a particular topic. Often, systematic reviews include a meta-analysis component which involves using statistical techniques to synthesize the data from several studies into a single quantitative estimate or summary effect size. Here is an excerpt (Section 1.1.2) from Higgins and Green (2011): "*What is a systematic review? A systematic review attempts to collate all empirical evidence that fits pre-specified eligibility criteria in order to answer a specific research question. It uses explicit, systematic methods that are selected with a view to minimizing bias, thus providing more reliable findings from which conclusions can be drawn and decisions made. The key characteristics of a systematic review are: a clearly stated set of objectives with pre-defined eligibility criteria for studies; an explicit, reproducible methodology; a systematic search that attempts to identify all studies that would meet the eligibility criteria; an assessment of the validity of the findings of the included studies, for example through the assessment of risk of bias; and a systematic presentation, and synthesis, of the characteristics and findings of the included studies*."

[51] The Crash Modification Factors (CMF) Clearinghouse is an online repository of CMFs, along with supporting documentation maintained by the FHWA.

[52] The evaluations in Arizona, Idaho, Illinois, Kansas, Ohio, Michigan, Texas (and perhaps elsewhere) were all disjoint efforts differing in method, variables and scope. The Stewardship report used data from a few states but the data was bad and the method unsound. The Bali et al. study that used data from ten states was a more concerted effort but still an isolated and inconclusive one.

what bits of evidence happened to materialize and then, by hook or crook, to mesh these into a number for the final table, an average thought to represent this unplanned harvest. There was no planned and coordinated effort to find out what the safety effect of edgelining really is.

The same unsystematic, piecemeal and designless approach to evidence creation persists. To illustrate, consider the relatively recent countermeasure of installing pedestrian countdown signals (PCS). The safety effect of PCS was studied in many places (San Francisco, San Jose, Detroit, Kalamazoo, Jacksonville, Gainesville, Toronto, Las Vegas, etc.). It was studied by different researchers using a variety of methods. Some found pedestrian crashes to have been halved, others found a modest change, some found that pedestrian crashes increased by 26%. This piecemeal approach to the conduct of evaluation research leaves the practitioner without evidence-base guidance; it allows decision makers the freedom to disregard research findings, and leaves the road user exposed to the harm of opinion-based practice.

It could perhaps be argued that there is no such thing as 'the safety effect of edgelining' or 'the safety effect of PCS', that there are large differences between states and cities, that the safety effect of treatments is always local, and that, therefore, a nationwide coordinated program of research makes no sense. Such an argument is difficult to sustain for long. First, because in spite of possible state-to-state and city-to-city differences there are also large commonalities. When humans drive similar vehicles on similar roads under similar traffic control regimes one may expect them to react to most treatments in a similar manner. Second, there is no natural limit to the local-effect argument. If the safety effect of some treatment in Michigan is thought to be markedly different from that in Arizona then should one not expect the safety effect of that treatment to also be significantly different in each of the 83 counties in Michigan or townships of the same county? Third, if the safety effect of everything was always local, there would be no use for meta-analyses or crash modification factors and one could not learn from the past experience of others. In short, there would be no rational road safety management. It seems to me that the conclusion is inescapable: while the safety effect of treatments may depend on some circumstances (variables) it is the task of research to determine which these variables are and what the dependence is. This is difficult to do by scavenging for bits of evidence produced by piecemeal, sporadic and uncoordinated research initiatives. Success requires conduct of planned research.

3.2.4 Issue 9: No plan, no framework.

When an action has important safety consequences research about these ought not to be limited by data that happened to materialize as a byproduct of local, sporadic and uncoordinated initiatives. To learn from experience there has to be a research plan and a framework for funding, producing and executing it[53].

[53] In the US the need for cooperation and coordination in research was recognized already in 1962 when the National Cooperative Highway Research Program (NCHRP) was set up. The role of the NCHRP is to "provide practical, ready-to-implement solutions to pressing problems" (http://www.trb.org/NCHRP/NCHRP Overview.aspx.) Money comes from state departments of transportation (The American Association of State Highway and Transportation Officials, AASHTO). Only research topics approved by at least two-thirds of the states are funded. While the NCHRP produced much valuable research it falls short of what is needed.
One limitation is the built-in dependence on State DOTs. Statements of research topics are solicited annually from AASHTO members, the chairs of AASHTO's committees and subcommittees, and the FHWA. What research is funded is decided by another AASHTO committee. By being represented on project panels, the DOTs can influence what the contractor will do, what is said in the final report, and whether and how reports are published. The potential for bias is evident.
Another limitation stems from being constrained to researching only pressing problems. Issues for which there are long-standing standards, warrants and practices, even if not evidence-based, are not likely to be thought 'pressing'. (Thus, e.g., it might be embarrassing to concede that determining the safety effect of edgelining was still a pressing problem in 2009.)

The role of evaluation research is not to be confined to dealing with 'pressing' problems; its role is to produce trustworthy evidence about the road safety consequences of policies and decisions that are likely to affect the safety of many road users.

Could there be a planned effort to find out what the safety effect of edgelining (or CPS) is? The answer to this question is yes. The Musick and Basile studies were too small to be conclusive but showed the way. There still are many two-lane rural roads that remain to be edgelined and, if someone really wanted to find out what the safety effect of edgelining is, it is feasible to conduct large scale randomized controlled experiments. Once carried out we will have defensible answers[54]. It is less a question of resources than of will. Of course the question can be asked not only about edgelining but about all design and operational decisions. I discuss these strategic issues in Hauer E, (2017a and b).

3.3 The second set of issues: Disregard of research findings

In the first anecdote Musick and Basile found that edgelining between access points seems to have harmed safety. One would think that DOTs and the professionals in their employ have an obligation to follow up on such indications[55]. But I did not find any sign of such; in both states edgelining between access points continued unabated. It is possible that the disturbing indications were thought a fluke, perhaps the conclusions were overlooked because they were clouded in jargon, or perhaps management did not read, trust, or chose not to heed what research found. Perhaps there was some reasonable explanation but because the actors are not around anymore I could not find it. However, based on what I can know, the disregard of research and the potential harm to road users is alarming.

Elsewhere in the same anecdote (footnote 18) I refer to sound research showing that placing raised pavement markers on two-lane rural roads is detrimental to safety. Still some states freely declared that they will not follow the guidance based on this evidence. Unlike the long-ago events of the first episode, this disregard of findings is contemporary. If an action is likely to harm road users and also costs money why do so[56]?

In the second anecdote I noted that in Idaho edgelining was associated with an increase in accidents on secondary roads and that the same happened in Illinois on narrow blacktop roads. I

These policies and decisions may be neither novel nor pressing. The production of trustworthy evidence requires planning and preparation by experts in evaluation research. Instead of scavenging for what data may be available one has to plan and prepare for collecting appropriate and sufficient data in conjunction with upcoming implementations.

[54]Salsburg (2001) relates the story of R.A Fisher's coming to the Rothamsted Agricultural Research Station early in the twentieth century. The results of ninety years of experimentation with 'artificial manures' (i.e. fertilizers) before Fisher's *arrival "was a mess of confusion and vast troves of unpublished and useless data ... The most that could be said of these artificial manures was that some of them worked sometimes, perhaps, or maybe."*(p.5-6) The Rothamsted scientists used various adjustments for differences in rainfall and soil fertility but Fisher showed that *"the year-to-year differences in weather and in 'artificial manures' were 'confounded' ... there was no way to pull them apart"*; this is reminiscent of the state of affairs in research on edgelining and on retroreflectivity.

For an experiment to lead to useful conclusions one has to be able to say that the effect that is measured is due to the treatment administered and not due to the nuisance influences over the change of which one has no control. This fact led Fisher to notion of randomization in the allocation of units to either 'treatment' or 'control'. The rest is history.

[55] The first and 'Fundamental Canon' in the code of ethics of the National Society of Professional Engineers (NSPE, 2007) says that in the fulfillment of their professional duties engineers must hold paramount the safety, health, and welfare of the public.

[56] RPMs make driving both easier and riskier. It is possible that the DOT officials thought that road users prefer driving comfort to risk to life and limb. Such a trade-off is both difficult to make and problematic to publicly (explicitly) defend.

did not find some follow-up research to ascertain that these troubling findings were incorrect. The findings were disregarded and both states continued to edgeline.

Also in the second anecdote the FHWA chose to denigrate and then to disregard the inconvenient finding that restoring the retroreflectivity of faded pavement markings was not associated with a reduction in crash frequency nor was the fading of pavement markings associated with an increase in crash frequency.

What makes it so easy for practice to disregard what research finds? This question is at the core of the relationship between research and practice. Perhaps, as discussed in Issue 3, in a culture where decisions affecting road-user safety are routinely based on opinion and judgment, research findings tend to carry little weight. Whatever the mix of reasons, the dysfunction is troubling. The 'unbearable lightness' with which research findings about the safety effect of actions can be disregarded in the design and operation of the transport system is one of the key symptoms of the ambiguous role of road safety research.

3.3.1 Issue 10: An Inquiry is needed

Data-based research is the foundation of evidence-based road safety management. Only by evidence-based management can one ensure that the road users' right to safety is safeguarded. If sound research is not done or is easily disregarded then a claim to evidence-based practice is vacuous. A high level inquiry into the role of road safety research is in order. Such an inquiry might illumine dark corners and raise troubling questions. Even so, since roads are a product the safety of which is shaped by the decisions made by public bodies and professionals in their employ, and because these decisions have safety consequence, practice should trust and heed findings of sound and applicable research.

3.4 The third set of issues: Perils of the funding-finding loop

In anecdote 2 the aim of the FHWA was to furnish objective-looking evidence to support a pre-determined course of action namely the introduction of minimum retroreflectivity requirements into the MUTCD. For this purpose it funded the writing of the 'Synthesis'. In the 'Synthesis' findings that favored the FHWA aim were made to look good even when based on bad data and methods and findings adverse to that aim were dismissed even when the research was sound. This illustrates what can happen when the findings of those who do the research are influenced by the agenda of those who disburse the funds.

An aspect the problem surfaced in a limited form in Section 2.5 where in-house research had to take into consideration the policy and practice of its parent agency. This led me to write about the importance of research independence (Issue 5). Another facet of the same peril led to the conclusion that evaluation and of execution need to be separate (Issue 6) because agencies that execute (the FHWA, the state DOTs and the large municipalities) do not like research findings that question their practice. Generalizing, it is the executing agencies that disburse the bulk of infrastructure-related road safety research funding and all researchers, not only those 'in-house', depend on their goodwill. This is the essence of the funding-finding peril.

The bending of research findings has undesirable consequences; here I will mention two. One is the creation of unfounded beliefs and thereby of inferior practice. The other and more damaging consequence is the harm to trust in research and thereby to its utility.

3.4.1 Unfounded beliefs

Miller (1991) was said to be one of the most cited papers on pavement marking safety; the 'Synthesis' (Carlson et al., 2009) must be similar in influence. Both were published in the Transportation Research Record, a publication favored by practitioners; both create unfounded beliefs. Had the refereeing been more rigorous the deficiencies of Miller's paper and the

tendentiousness of the 'Synthesis' might have been noted. But, as they were not, the content of both papers continues to influence what practitioners believe. Practitioners are led to believe that edgelining benefits safety in all circumstances, that the benefit is a matter of consensus and need not be questioned, that the benefit is large and well substantiated, that edgelining does not affect speed, that increasing retroreflectivity is always good for safety etc. Some of these beliefs are incorrect, others just unfounded. Evidence-based practice cannot be based on beliefs that are untrue or unfounded and practice that is not evidence-based is likely to waste resources, life and limb, or both. While it is impossible to keep the reservoir of knowledge free of impurities, one can reduce their incidence and harm. The means for doing so are at least four:

Quality refereeing[57]. By recommending publication the referee vouches for the quality of findings. In safety evaluation research this cannot be done without access to the data and without sufficient time (resources) to examine the path between the data and the conclusions.

Conflict of interest declaration. Research findings can be colored by self-interest. The sources of self-interest are many: the need to please employers and organizations, the desire to obtain research current and future contracts, etc. The duty to comprehensively declare such conflict should be mandatory not only in journal publications but in all reports.

Encouragement of responsible criticism[58]. The North American professional culture does not encourage criticism. And yet, without criticism, unsound findings will go unexposed. Bad findings reduce the value of good work. Responsible criticism has to be brought about by pre-meditated means. One way is by providing for an audit of the data and its analysis[59]; another is by institutionalizing the production of systematic reviews with a chapter devoted to a critical review of the evidence used.

Replication. Evaluation research is about actions that have important consequences. One should therefore have confidence in the findings. An important confidence building measure is replication. Findings that cannot be verified by replication cannot be used responsibly.

3.4.2 Issue 11: Guarding against unfounded beliefs

The aforementioned means for guarding the reservoir of knowledge against impurities require structured action. Thus, to referee and re-examine data-based research findings there has to be a maintained repository of data and funding for independent re-analyses. To mandate conflict of interest statements requires agreement on what must be declared and resolutions to make these binding. To have systematic reviews that include a critical review of past results there has to be a planned and funded program for their production. The need for to replicate findings has to be acknowledged, planned for, and required.

That the funding-finding loop creates unfounded beliefs and thereby of inferior practice is bad enough. However, even more damaging consequence is the harm to the trustworthiness of research and thereby to its utility.

[57]The shortcomings of the refereeing process for the Transportation Research Record are known but not easy to correct without substantive changes to the annual Transportation Research Board meeting.

[58]Even though it is possible to submit a discussion for a paper to be published in the Transportation Research Record, one rarely sees one.

[59] Most of the money and effort in evaluation research goes into the assembly and preparation of data. And yet, astoundingly, very seldom is the costly data set available for re-analysis by other researchers.

3.4.3 Harm to trust

We live in an era of disillusionment with government and elites. At the core of this crisis is the suspicion that what we are told serves someone else's interest, not ours[60]. Because the road system is produced and operated by governments and these also initiate and fund research, the same skepticism extends to the findings of evaluation research in road safety. As the anecdotes show, it is occasionally justified.

When research findings are not widely trusted then facts become fluid, motives suspect, unfounded assertions abound, and few can be shown wrong. In such an atmosphere non-political decision-making is without its source of legitimacy. As a result the decisions needed for the functioning of a complex society will increasingly be made with less regard for relevant facts. I view this trend with alarm. The only way to stem and reverse it is by ensuring that the aim of research[61] is to find out 'what is' and not what someone 'would like it to be'. The distinction is important, due to the prevalence of the latter circumstance. To illustrate, the Ohio and Kansas DOTs would have liked edgelining to be beneficial always and everywhere; the FHWA would have liked to be able to show that when retroreflectivity drops below some minimum then accident frequency increases; the American Glass Bead Manufacturers Association would have liked edgelining to show a large benefit-cost ratio even on low volume roads, etc. Having studied the reports produced in the aforementioned circumstances (and in many similar ones) I could recognize the link between what was reported and what someone wished to be found. But not many have the resources to do so. And thus the corpus of research findings is an undifferentiated agglomeration of 'what is' and what some 'would like it to be'.

The job is to separate the wheat and the chaff, to promote trustworthiness, and to reduce nihilism. Doing so is important to both the left and the right on the political spectrum. It is important to those who believe in government intervention because their opinion is defensible only if the government can get it mostly right. In the Garden of Eden one could tell right from wrong by biting an apple, in modern society you need the institution of independent research. It is important to those on the right who wish to diminish the role of government and for the same reason. Only independent research can tell without bias what the costs and benefits of programs and interventions are. The insulation of research from agendas is important also to those who do research and those who make use of it. It is a personality trait of most researchers that they aim to discover what is true. When this does not happen to suit the perceived needs of those who ordered the work, the conflict is corrosive. A similarly corrosive situation may afflict users of research when findings conflict with the perceived interest of higher-ups. The confluence of all these interests may bode well for the resolve to insulate research from agendas.

The need to separate church from state and the legislative from the executive power has been recognized a long time ago. It is time to consider the need to separate profession-serving evaluation research from the executive institutions. To furnish a catchy phrase, one has to disrupt and restructure the 'Funding-Finding' loop. In this loop linkage goes in both directions. In one direction, those who do the funding and have interests that may influence what gets reported as findings. Going in the other direction, the future livelihood of those who do the 'finding' (universities, research institutes, consultants, etc.) depends on those who do the funding.

---

[60] For example, "Onora O'Neill has ascribed this lack of trust to medical institutions and professionals introducing measures that benefit themselves, not the patient." https://en.wikipedia.org/wiki/Autonomy.

[61] I examined several definitions of the word 'research'. The definition I like most is "*the systematic investigation into and study of materials and sources in order to establish facts and reach new conclusions*". I like it because it speaks about establishing facts and mention that the conclusions must be new.

3.4.4 Issue 12: Insulate research from extraneous interests

Because almost all road safety research funding comes from the 'State' the existence of the entire research industry depends on its goodwill. The need to disrupt and restructure this master-servant loop is evident. Unfortunately, the actors within the loop cannot bring such change about. Perhaps the National Academies of Sciences, Engineering and Medicine has the necessary constitution, authority and impartiality.

4 CLOSURE

This paper is about the relationship between road safety research and the practice by which the road infrastructure is built and operated. The current relationship evolved over time in a piecemeal manner and was shaped by the practicalities of the day and the interests of the main participants. But one of the main participants – the road user whose life and limb is at stake – is represented only by proxy. Road users have to trust that those who produce and operate roads act on factual information and safeguard their interest. This, as I have shown, does not always work as it should.

I examined the functioning of the research-practice relationship in the light of two historical anecdotes. In these, so it seems, it was easy to supplant opinion for factual information and to make research subservient to other considerations. The interest of the road-user and the taxpayer was not well served. These anecdotes allowed me to point out several dysfunctions and issues of concern. To the extent that similar issues and dysfunctions are common, action is required.

Just describing some dysfunctions of the current research-practice relationship, as I did here, is not likely to induce action. Too many have too much at stake in the status quo. Action is likely to come when road-users and their political representative will recognize that those building and operating roads do not rely on a well-functioning research function to base their practice on.

Progress has been made towards evidence-based practice; research and researchers enabled this progress. But much of practice is still opinion-based and the role of research ambiguous. The first step towards reform is rethink and then revamp the research-practice relationship. The reformed relationship should be endowed with a purposeful structure, one that cures what dysfunction there is and promotes generation of trustworthy information.

REFERENCES

AASHTO, (2010). The Highway Safety Manual, First Edition. American Association of State Highway Transportation Professionals, Washington, D.C.,

AASHTO, (2011). A Policy on Geometric Design of Highways and Streets, 6th Edition.

Agent KR, N Stamatiadis, and S Jones, (1996). Development of Accident Reduction Factors. KTC-96-13, Kentucky Transportation Cabinet.

Bahar G, Mollett C, Presaud B, Lyon C, Smiley A, Smahel T, and McGee H, (2004). Safety evaluation of permanent raised pavement markers. NCHRP Report 518. Transportation Research Board.

Bahar G, M Masliah, T Erwin, E Tan, E Hauer, (2006). Pavement Marking Materials and Markers: Real-World Relationship Between Retroreflectivity and Safety Over Time, NCHRP Web Only Report 92, can be viewed at the following Internet Web site: h*ttp://onlinepubs.trb.org/onlinepubs/nchrp/nchrp_webdoc_92.pdf.*

Bahar G, M Masliah, R Wolff, and P Park (2007)*. Desktop Reference for Crash Reduction Factors.* Publication FHWA-SA-07-015. Federal Highway Administration, U.S. Department of Transportation, Washington, D.C.

Bali S, R. Potts, JA. Fee, JI Taylor, and J Glennon, (1978). *Cost-Effectiveness and Safety of Alternative Roadway Delineation Treatments for Rural Two-Lane Highways.* Report FHWA-RD-78-50. National Technical Information Service, Springfield, Va.

Basile, A. J. (1962). Effect of Pavement Edge Markings on Traffic Accidents in Kansas. *Highway Research Board Bulletin 308*, 1962, pp. 80-86.

Benessia, A., Funtowicz, S., Giampietro, M., Guimarães Pereira, A., Ravetz, J., Saltelli, Strand, R., van der Sluijs, J., 2016, The Rightful Place of Science: Science on the Verge, Published by The Consortium for Science, Policy and Outcomes at Arizona State University.

Botha JL, Zabyshny AA, Day JE (2002). Pedestrian Countdown Signals: An Experimental Evaluation. Vol. 1, San Jose, CA.

Carey MP (2013). The Federal Rulemaking Process: An Overview. Congressional Research Service**,** www.crs.gov, RL32240.

Carey MP (2014). Cost-Benefit and Other Analysis Requirements in the Rulemaking Process. Congressional Research Service**.** www.crs.gov. R41974.

Carlson PJ, ES Park, CK Andersen, Carl K (2009). Benefits of Pavement Markings: A Renewed Perspective Based on Recent and Ongoing Research. Transportation Research Record 2107, Journal of the Transportation Research Board, pp 59-68 .ISSN: 0361-1981. Downloaded from http://safety.fhwa.dot.gov/roadway_dept/night_visib/pavement_visib/no090488/

Carlson PJ, ES Park, and DH Kang. (2013). Investigation of Longitudinal Pavement Marking Retroreflectivity and Safety. Transportation Research Record: Journal of the Transportation Research Board, No. 2337, Transportation Research Board of the National Academies, Washington, D.C., pp. 59–66.

Carlson PJ, RE Avelar, ES Park and D Kang, (2015). Nighttime Safety And Pavement Marking Retroreflectivity On Two-Lane Highways: Revisited With North Carolina Data. Presented at the 94th Annual Meeting of the Transportation Research Board. Paper 15-5753.

Chemerinsky E, (2001). Against Sovereign Immunity. Stanford Law Review. 53, 1221-1224.

Congress (1992). Section 406 of the Department of Transportation and Related Agencies Appropriations Act, 1993 (Pub. L. 102–388; October 6, 1992).

Donnell ET, V Karwa, Sr Sathyanarayanan (2009). Analysis of the effects of pavement marking retroreflectivity on traffic crash frequency on highways in North Carolina. Transportation Research Record 2103, pages 50-60.

Elvik, R., and T. Vaa. (2004) The Handbook of Road Safety Measures*.* Elsevier, Amsterdam, The Netherlands.

Executive Order 12866, Regulatory Planning and Review, Federal Register 58 , No 190. Presidential Documents. October 4, 1993. http://www.whitehouse.gov/omb/ inforeg/ eo12866.pdf.

FHWA (2009). The Manual on Uniform Traffic Control Devices.

Federal Register, Vol. 75, No. 77, April 22 2010/Proposed Rules, 20935-20941.

Fosser, S*.* (1992) An Experimental Evaluation of the Effects of Periodic Motor Vehicle Inspection on Accident Rates. *Accident Analysis &Prevention,* Vol. 24, No. 6, pp. 599-612.

Gan A, J. Shen, and A Rodriguez (2005). Update of Florida Crash Reduction Factors and Countermeasures to Improve the Development of District Safety Improvement Projects. Florida Department of Transportation.

Halverson, JR (1957). Right Shoulder Guide Lines, Traffic Safety, Chicago, IL, No. 3, pp. 20-22.

Hauer, E., 2000a. Safety in Geometric Design Standards I: Three Anecdotes. Proceedings of the 2nd International Symposium of Highway Geometric Design, R. Krammes and W. Brillon (eds.). Forshungsgeselschaft für Strassen und Verkehrsvesen e.V., Köln, pp. 11-23. Downloadable from ResearchGate.

Hauer, E., 2000b. Safety in Geometric Design Standards II: Rift, Roots and Reform. Proceedings of the 2nd International Symposium of Highway Geometric Design, R. Krammes and W. Brillon (Eds.). Forshungsgeselschaft für Strassen und Verkehrsvesen e.V., Köln, pp. 24-35. Downloadable from ResearchGate.

Hauer, E, (2007), A Case for Evidence-Based Road-Safety Delivery. Published in "Improving Traffic Safety Culture in the United States. The Journey Forward". AAA Foundation for Traffic Safety. April 2007. https://www.aaafoundation.org/sites/default/files/SafetyCultureReport.pdf.

Hauer E, (2016). An exemplum and its road safety morals. Accident Analysis and Prevention.94, 168–179.

Hauer E, (2017a). Developing Consensus in Research about the Safety Effect of Manipulations. Appendix G of CarterD, R Srinivasan, F Gross, S Himes, T Le, B Persaud, C Lyon, and J Bonneson, (2017) Guidance for the Development and Application of Crash Modification Factors, National Cooperative Highway Research Program, Project 17-63, Final Report, Submitted March 2017 (unpublished).

Hauer E, (2017b). Enhancing Future CMF Research. Appendix L of Guidance for the Development and Application of Crash Modification Factors, NCHRP Project 17-63.

Hearings before the Subcommittee on Surface Transportation of the Committee on Public Works and Transportation. House of Representatives. One Hundred Second Congress. U.S. Government Printing Office, 1991.

Higgins JPT and Green S, eds., (2011). Cochrane Handbook for Systematic Reviews of Interventions. Version 5.1.0. [Updated March 2011].

ITF (2016). *Zero Road Deaths and Serious Injuries: Leading a Paradigm Shift to a Safe System*, OECD Publishing, Paris. DOI: http://dx.doi.org/10.1787/9789282108055-en.

Jackson J (1983). Reflectorised Road Markings: Their Contribution to Traffic Safety. *Proc., Symposium on Visual Aspects of Road Markings,* Paris, France.

Kallberg, V.P. (1993). Reflector posts—Signs of danger? *Transportation Research Record, 1403*, 57-66.

Kapoor S and Magesan A (2014). Paging Inspector Sands: The Costs of Public Information. American Economic Journal: Economic Policy, vol. 6, no. 1, February 2014, pp. 92-113.

Karwa V, AB. Slavkovic and ET Donnell (2011), Causal inference in transportation safety studies: Comparison of potential outcomes and causal diagrams. The Annals of Applied Statistics. Vol. 5, No. 2B, 1428–1455.

Lakey RT, (2016) Keep Science and Scientists Credible: Avoid Stealth Policy Advocacy. Bulletin of the Ecological Society of Australia. 46 (3), 14-15.

Macy PR, ND Nedas, and GP Balcar (1980) or Potters Industries Inc. (1981). Road Markings as an Alcohol Countermeasure in Traffic Safety: A Field Test of Standard and Wide Edgelines. Final Report to FHWA.

Markowitz T, Sciortino S, Fleck JL and Yee BM (2006). Pedestrian countdown signals: Experience with an extensive pilot installation. ITE Journal, 76, 1, 43-47.

Marsh PA, (1973). Evaluation of Minor Improvements - Edge Striping, Department of Highways, Traffic Section, Boise, Idaho.

Michigan Department of Highways (?).Pavement Edge Markings Reduce Accident Rates", Hi-Lighter,I, No. 5, pp. 12-13

Miller T.(1991). Benefit-Cost Analysis of Lane Marking. In *Transportation Research Record 1334*, Transportation Research Board, National Research Council, Washington, D.C., 1991, pp. 38-45.

Mohan D (2003). Advancing Safety as a Fundamental Human Right. Health and Human Rights Journal, Vol. 6, No. 2. Pp. 161-166.

Mulvaney CA, Smith S, Watson MC, Parkin J, Coupland C, Miller P, Kendrick D, McClintock H. (10 December 2015).Cycling infrastructure (changes to the road environment) for reducing cycling injuries in cyclists. Downloaded on 3/21/2017 from http://www.cochrane.org/CD010415/INJ_cycling-infrastructure-changes-road-environment-reducing-cycling-injuries-cyclists

Musick, J. V. (1960). Effect of Pavement Edge Marking on Two-Lane Rural State Highways in Ohio. Highway Research Board Bulletin 266, pp. 1-7.)

Nedas ND, Balcar GP, and Macy BP (1980). Road Markings as an Alcohol Countermeasure forHighway Safety: Field Study of Standard and Wide Edgelines. Transportation Research Record 847.Pp. 43-46.

NSPE (National Society of Professional Engineers), 2007, Code of Ethics. Publication #1102.

Office of Highway Safety (1981).The 1981 Highway Safety Stewardship Report. Report of the Secretary of Transportation to the United States Congress, FHWA, U.S. DOT Washington, D.C.

O'Neill, Onora (2002). Autonomy and Trust in Bioethics (The 2001 Gifford Lectures.) Cambridge University Press.

Owsley C and McGwin Jr. G (2010). Vision and Driving. Vision Research, 50(23): 2348–2361.

Pulugurtha SS, Desai A and Pulugurtha NM (2010). Are Pedestrian Countdown Signals Effective in Reducing Crashes? 3Traffic Injury Prevention, 11, 632–641.

Richmond SA, AR Willan, L Rothman, A Camden, R Buliung, C Macarthur, A Howard, (2013). The impact of pedestrian countdown signals on pedestrian-motor vehicle collisions: a reanalysis of data from a quasi-experimental study. Injury Prevention, Volume 20, Issue 3, *155-158*

Rudin-Brown CM and SL Jamson, ed. (2013). Behavioural adaptation and road safety. CRC Press.

Salsburg D (2001). The lady tasting tea, Henry Holt and Company, New York.

Sharfi T and Shinar D, (2014). Enhancement of road delineation can reduce safety. Journal of Safety Research,49 pages 61-68.

Smadi O, N Hawkins, I Nienanya, B Aldemir-Bektas (2010), Pavement Markings and Safety.Center for Transportation Research and Education. Iowa State University. HRB Project TR-580.

Smiley A, (2008). Driver behavior: A moving target. TR News 254, pp. 19-24.

State of Illinois Department of Public Works and Buildings (1959). Traffic Accident Experience Before and After Pavement Edge Line Painting. Division of Highways, Bureau of Traffic, Springfield, IL.

Statement of the American Traffic Safety Association to the Subcommittee on Surface Transportation. Hearing on Highway Safety Issues. April 18, 1991. Pages 1501-1504 of the

Hearings before the Subcommittee on Surface Transportation of the Committee on Public Works and Transportation. House of Representatives. One Hundred Second Congress. U.S. Government Printing Office, 1991.

Stock, JR., JK Weaver, HW Ray, JR Brink, and MG Sadof.(1983). Evaluation of Safe Performance Secondary School Driver Education Curriculum Demonstration Project. Washington, DC: U.S. Department of Transportation, National Highway Traffic Safety Administration.

Sun X, and VO Tekell. (2005). Impact of Edge Lines on Safety of Rural Two-Lane Highways.LTRC Report 414. Louisiana Transportation Research Center, Baton Rouge, Louisiana.

Tamburri TN, CJ Hammer Jr., JC Glennon and A Lew, (1968), Evaluation of minor improvements. Highway Research Record 257, 34-79.

Transportation Research Board (2008) .Safety Research on Highway Infrastructure and Operations.Improving Priorities, Coordination, and Quality. Special Report 292.

Tsyganov, AR, RB Machemehl, NM Warrenchuk, and Y Wany (2006). *Before-After Comparison of Edge Line Effects on Rural Two-Lane Highways.* CTR Technical Report 0-5090-2. Center for Transportation Research, The University of Texas, Austin, Texas.

Van Driel, CJG, RJ Davidse, and MFAM van Maarseveen. (2004) The Effects of an Edge Line on Speed and Lateral Position: A Meta-analysis. *Accident Analysis and Prevention,* Vol. 36, No. 4, pp. 671-682.

Wasserstein RL & Nicole A. Lazar (2016) The ASA's Statement on p-Values: Context, Process, and Purpose, The American Statistician, 70:2, 129-133, DOI:10.1080/00031305.2016.1154108

Willey, WE (1955). Arizona's Dashed Shoulder Stripe, Traffic Quarterly, Saugutauck, CT, IX, No. 2,pp. 212-219.

Yi Jiang, (2006). The Effectiveness and Criteria for Placement of Raised Pavement Markers (Synthesis Study). FHWA/IN/JTRP-2006/6. Purdue University, Joint Transportation Research Program, Project No. C-36-59S.

APPENDIX A. THE 'SYNTHESIS' ON THE SAFETY EFFECT OF EDGELINING

The 'Synthesis' begins the review of edgelining (on pages 5-7) with the Musick and Basile studies. It repeats only the positive messages which Musick and Basile, in their circumstance, thought appropriate to include in their summaries. The 'Synthesis' does not mention that, contrary to expectation, edgelining was found harmful in-between access points, that it affected not only night-time but also daytime accidents, that in Kansas it was associated with an increase in nighttime accidents etc.[62].Doing so would not support the rulemaking which was to show that it is the improved visibility of pavement markings between access points and at night which (due to more retroreflectivity) helps drivers to be safe.

Next the 'Synthesis' reviews the findings of the 'Stewardship Report' (Office of Highway Safety, 1981) calling it *"one of the most robust studies"*. It is said to have found that *"Adding edge lines resulted in a statistically significant 16 percent decrease in nighttime crashes …"* (Synthesis, page 5). Alas, the 'Stewardship Report' is neither robust nor does it serve the mission of the 'Synthesis'. To understand why, context is relevant.

By the Highway Safety Act of 1973 Congress provided federal funding for pavement markings. The 'Stewardship Report' is an accounting to Congress by the Secretary of Transportation about the effect to which the money was spent. It would not look good if accident savings proved to be insufficient. It was therefore disconcerting that when the FHWA examined the State data the simple ("gross") measures of safety benefit were disappointing[63]. Naturally this was of concern and *"… to address some of these concerns a special pavement marking evaluation was conducted"*(Stewardship Report, Page 51). Unfortunately again, this special evaluation rested on

[62]Thus, e.g., on page 5 the 'Synthesis' reports *"a 63 percent reduction in crashes at access points"* but does not say that away from access points there was an increase. (In Ohio the increase was of15% and in Kansas of 27% while in Ohio there were about two times more accident away from access points than at access points and in Kansas the ratio was close to three.) The 'Synthesis' repeats uncritically the unreasonable claim that *edgelining "resulted in a 78 percent reduction in fatalities"* (page 5), a claim based on the counts from Basile's Table 4 and shown below. The 78% reduction comes from assuming that were the edgelined section not edgelined one should expect 5×15/4=18.75 fatalities, and that because 4/18.75=0.21 it is a 79% reduction. (The difference between 78% and 79% is due to rounding).

From Basile (1962), Table 4		Before	After
Treated (edgelined)	Killed	5	4
	Killed & Injured	110	104
Control (not edgelined)	Killed	4	15
	Killed & Injured	140	115

But note that that while the number of persons killed on the control sections increased from 4 to 15 the number of persons injured there decreased from 136 to 100. A more sensible approach would have been to use the 'Killed & Injured' numbers for 'control'. With this assumption, were the edgelined section not edgelined one should expect 5×115/140=4.1 fatalities, about the same number as what was observed to occur. In this, as in many other cases, what is reported depends what the researcher assumes, often without mention and good reason.

[63]On those roads where only edgelines were added the number of fatal accidents increased by 4% and the number of injury accidents increased by 10% (Table III-13). When miles of travel were taken into account the fatal accident rate declined by 5% but injury accident rate increased by 7% (Table III-14). Trying to explain away this inconvenient result the authors of the '1981 Stewardship Report' speculated (on page 51) that if during the 'before edgelining' period many of the edgelined roads had an unusually low accident rate then, just by regression-to-the-mean, the accident rate would have tended to increase in the after-edgelining period. This argument does not hold water well. I see no reason why practitioners would choose to edgeline low accident rate roads; doing the opposite makes more sense. Be it as it may, whether the speculation had a basis in fact could have been checked against data but, within the report, there is no sign that this was done.

bad data[64] and on a bad assumption. The 'special evaluation' "… *assumed that pavement markings have little or no effect on reducing day accidents…* (Page 51)." As shown in the footnote when the assumption is not true the findings are different[65]. Thus, the validity of the conclusion (that there was a 16% decrease in nighttime crashes) depends on an unsupported assumption. The verity of this core assumption was not checked against data or extant research findings[66], and was contradicted by studies about how edgelining affect vehicle speed and placement, information already available at that time. Results based on an assumption that is contradicted by available empirical evidence do not amount to 'findings' and certainly are not 'robust'.

However, the problem is not only that the data was bad and the core assumption demonstrably untrue; the problem is also one of basic logic. You may not make an assumption from which it follows by logic that what you wish for is true. The Synthesis aimed to show that by enhancing pavement marking retroreflectivity one will enhance safety at night. The 'no-effect-on-daytime-accidents' assumption logically implies that whatever improvement in safety an empirical study will find, it is only at night. Even if the safety improvement was entirely during daytime, once the no-effect-on-daytime accidents assumption is made it will appear that the improvement is only at night. In this manner, as long as there was an accident reduction, confirming what the Synthesis hoped for is a forgone conclusion.

Still on edgelining, the 'Synthesis' devotes a section to 'More Recent Pavement Marking Safety Studies'. The Federal Register relies on these because these results are allegedly based on more

[64] The '1982 Stewardship Report' says (on page 49) that what complicates the evaluation was that *"Accident data files are becoming more and more complete which, in part, could account for increases in the number of accidents recorded during the evaluation 'after' period. In Minnesota, for example, the number of reported 'after' accidents more than doubled principally because of the increased ability to locate and report accidents… ."* Even so, two pages later in Table III-17, the same report lists a 72% statistically significant decline in injury accidents; this in spite of the fact that the number of accidents in the control group increased inexplicably from 399 in the 'before' period to 967 in the 'after' period of same duration. Bad data do not lead to good conclusions. Subjecting bad data to a test of statistical significance, as has been done, can only make for statistically significant nonsense.

[65] Suppose that the no-effect-on-daytime-accidents assumption was not quite right;that the presence of an edgeline caused drivers to increase their daytime speed by, say, 2% and thereby increased the reported daytime injury accidents by about 8%. This relationship between mean speed and injury accidents is roughly in accord with what research shows. Consider now the following invented data:

	Before	After
Daytime	100	(a)100, if edgelining had no effect on daytime accidents (b)108, if edgelining increased daytime accidents by 8%
Nighttime	80	73

Had edgelining had no effect on daytime accidents (case a) one would estimate that nighttime accidents were reduced by 8.8% (73/80=0.912). Had edgelining increased daytime accidents by 8% (case b) and if one assumed that it did not, one would expect without edgelining 80×108/100=86.4 nighttime accidents in the 'after' period. Now the nighttime accident would seem to have been reduces by 15.5% (73/86.4=0.845).What in truth was wash (+8% during the day -8.8% at night) would now seem to be substantial net benefit (0% assumed during daytime and -15.5% at night).

[66] Thus, e.g., in 1959 the State of Illinois study found that following edgelining accidents in daytime declined more than nighttime accidents.

data and better techniques of analysis[67]. In this section the 'Synthesis' reviews the work of Miller (1991) and mentions the meta analyses by Elvik and Vaa (2004) and Bahar et al. (2007).

About Miller (1991)[68] the 'Synthesis' says that it is "*one of the most often cited pavement marking safety studies"* and that "*… using studies deemed credible* (Miller found that*) an average crash reduction of 21 percent could be attributed to pavement markings*."(Page 6.) Are these studies credible? It is instructive to check.

Miller lists (in Table 4 on page 4) nine estimates of what others found to be the safety effect of edgelining in North America. The first two are from Musick and Basile reported in 1962. They were introduced into evidence in the 'Early Pavement Marking Safety Studies' section of the 'Synthesis', and I reviewed them in the first anecdote where their strengths and limitations were noted.

The third estimate listed by Miller comes from Bali et al. (1978)[69].As I show below, what Miller extracts from Bali et al. (and makes into the kingpin of his cost-benefit analysis) is a self-contradictory improvisation. Bali at al. collected data at 500 sites in ten states and described the results in thoughtful detail. Their main conclusions about the safety effect of edgelining come from comparing the accident rates of sites with and sites without edgelines. In the executive summary Bali et al. say that *"Results of analyses of accident rates at sites with edgelines versus those without edgelines were mixed."*(Page 10). In the final report they say that *"Edgelines were found to have little or no effect on roadway accidents."* (Page 51). In addition to the aforementioned cross-section comparisons Bali et al. fitted seven stepwise regression models to their data. Edgelining shows up as an explanatory variable in four out of the seven regression equations. In three of the four the regression coefficient is positive meaning that edgelined roads have a higher accident rate than roads without edgelines. Bali et al. also report on their limited success with before-after edgelining comparisons. In about half of these (Table 13, page 62) the addition of edgelines was associated with an increase in accident rate.

How then could Miller claim (in his Table 4, Page 42) that Bali et al. found that ("Nationwide") edgelining reduces accidents by 8%? I did not find this number in the Bali et al. report. Miller must have extracted it from their Table 16 (on page 87) where Bali et al. give average accident rates for sites in seven different delineation treatment classes. Thus, e.g., they find that sites without edgelines where the centerline had only raised pavement markers had an average of 1.67 accidents per million vehicle miles (MVMs). On sites without edgelines where the centerline had raised pavement markers and may have also been painted the average accident rate was 2.12 accidents/MVM. Sites with any centerline and also edgelines had an accident rate of 1.96 accidents/MVM. Using the ratio 1.96/2.12=0.92 Miller must have concluded that the addition of edgelines to some kind of centerline caused an 8% reduction in the accident rate. Note that from the same numbers that Bali et al provide Miller should have concluded that because 1.67 is less than 1.96, it is safer not to have edgelines at all, just raised pavement markers on the centerline.

Estimates of this kind when based on the comparison of raw accidents rates from cross-section data are commonly regarded as confounded and unreliable. Bali et al. knew that and cautioned*: "It is important to note that the accident rates given* (in their Table 16) *are for all sites within a given treatment/situation class. It may be then, that the differences in accident rates are attributable to differences in climate, traffic volume, roadway width, etc., associated with particular delineation/situation combinations, rather than solely to changes in delineation."* (Page 90). But

---

[67] In the Federal Register (Page 20936) the FHWA says that *"While early landmark studies are referenced in the synthesis, the emphasis was directed to more recent studies offering new insights into the benefits of pavement markings that were previously undetectable (more data are now available for advanced analysis techniques.)"*

[68]Miller's paper (1991) was paid for by the American Glass Bead Manufacturers Association.

[69] Miller says (on page 41) that this is *"The best U.S. effectiveness study …"*

Miller did not heed this caution. In spite of what Bali et al. repeatedly say about their own results[70] Miller extracted the 8% estimate from two disparate raw accident rates pertaining to different roads and based his cost/benefit analysis for the edgelining of rural two-lane roads on this flimsy foundation[71].

Five of Miller's estimates in Table 4 were copied from Potters (1981)[72]. But Potters did not produce these estimates; they merely extracted them from still earlier research. After some digging I obtained copies of four of those ancient reports.

One of these is about experience in Arizona where, according to Potters, edgelining reduced accidents by 60%. The reference is to Wiley (1955) and the data is for two roads, each with 15 miles edgelined and 15 miles left without edgelines. There was a grand total of 53 accidents and therefore Wiley could afford to list in the paper some detail about each. Some occurred during day some at night, some involved one vehicle some several, in some accidents the driver was asleep, in others the collision was with a horse, or the tire blew out, or the car was on fire, or improperly parked, or the wheel was faulty, or the vehicle made an improper turn, etc. Wiley does not mention the 60% reduction. As best I can make out it comes from counting all accidents that occurred on the edgelined 30 miles and comparing them to those on the un-edgelined 30 miles. About the road that are being compared we only know that their paved surface was 40' wide. We do not know how many access points of what kind were on the edgelined and un-edgelined halves or how they might have differed in the other relevant aspects. In short this is closer to an old professional anecdote than to a research study.

The second report which Potters list and Miller copies is by the State of Illinois (1959) where edgelining is said to have reduced accidents by 21%. (There were 683 accidents one year before edgelining and 540 one year after edgelining.) Changes in the amount of traffic, secular trend, or possible regression to the mean were not accounted for. Some findings were said to be *"opposite to expectation."*(Page 3). Thus, e.g., daytime accidents declined more than nighttime accidents, non-collision accidents and accidents due to being on the other side of the road increased, accidents on concrete pavement declined but on narrow blacktop increased.

The third estimate which Potters list is a 15.9% accident reduction due to edgelining for Idaho (Marsh, 1973)[73]. The Idaho study was also a naïve before-after study similar to the State of Illinois one, except that here the threat of the regression-to-mean-bias is explicit and evident. On page 2 Marsh says *that "Sections of highway were designated for edge striping because of past accident record."* If so the count of accidents was expected to have dropped from 'before' to 'after' even had the roads remained without edgelining. As in Illinois, the Idaho estimates are also all over the place. Thus, e.g., on the primary road system Marsh finds an accident reduction of 24% but on the

[70] That the results are mixed, that edgelines are of no or little effect, that in most regressions edgelining seems to increase the accident rate, and that what differences in accident rate exists may be due to many unaccounted-for factors.

[71] In his 'Conclusions' section Miller adds that *"…Much of the safety benefit is achieved during periods of poor visibility. That suggests checking roadway retroreflectivity regularly and restriping promptly when retroreflectivity drops below recommended levels."* (Page 45). Conclusions usually reflect something that has been shown within the paper. But in the body of Miller's paper nothing is said to support the 'low visibility' and 'retroreflectivity' claims present in his conclusion. It is an unfounded afterthought which is in line with what the president of ATSSTA said to Congress in the same year (see Statement, 1991). It also must have pleased the American Glass Beads Manufacturers Association who are affiliated with ATSSA and who paid for Miller's study.

[72] I have a report by Potters Industries (1981) with the same title as Macy et al (1980). The Potters report provides the references to these five studies listed by Miller and attributed by him to Macy et al. (1980). Potters Industries is a leading producer of engineered glass materials.

[73] I did not find in Marsh a trace of the 15.9% reduction listed in Potters.

secondary system there was an increase of 14.7%. On secondary roads daytime accidents increased by 50% while nighttime accidents decreased by 20.8% but on the interstates daytime accidents diminished by 39.7% while nighttime accidents increased by 12.1%.

The fourth estimate listed in Potters is for Utah, where edgelining is said to have caused a 38% accident reduction. The reference is to Halverson (1957). But in the reference I have there is no such estimate. Halverson only describes an example of a road on which during 19 months before edgelining there were 2 accidents and during 9 months after edgelining there were 8 accidents and says (on page 19) that this is *"encouraging"*. Halverson acknowledges that *"Our accident studies have been limited, with samples too small to be of statistical significance. However, common sense tells us that the guide lines have been of great help to motorists at night. Our feeling on this matter has been reinforced by newspaper editorials and the many letters received complimenting the department on the pavement edge markings."* (Pages 20-21)

Miller copied the 38% reduction for Utah that is listed in Potters and indicated, in addition, that the same 38% reduction is supported by Jackson (1983)[74]. After some more archeology I found that the Jackson (1983) paper contains no data, no safety effect estimates, and no reference to Utah.

The last North American estimate listed in the Potters report and copied by Miller is a 3% reduction for Michigan. Here the reference is to a two-page article by the Michigan Department of Highways which I could not unearth anymore.

Thus, on closer examination, there is nothing recent about the North American evidence in Miller. In fact all the estimates reviewed predate the '1981 Stewardship Report' and therefore Miller's paper should not be in the 'More Recent Pavement Marking Safety Studies' section of the 'Synthesis'. More importantly, the Federal Register says (on page 20936)that in the 'Synthesis' "*emphasis was directed to more recent studies offering new insights into the benefits of pavement markings that were previously undetectable (more data are now available for advanced analysis techniques.)"* Not only are these studies not recent, there is nothing about them that can be characterized as using *"more data or better techniques of analysis"*.

Also mentioned in the "More Recent Pavement Marking Safety Studies" section of the 'Synthesis' are the meta-analyses of Elvik and Vaa (2004) and Bahar et al. (2007). Elvik provided me with the list of North American references on the safety effect of edgelining that were used; these are Musick(1980), Basile (1962) and Tamburri et al. (1968)[75].Thus, once again, the references are neither recent nor based on more data and improved method. The Bahar at al. estimates of the safety effect of edgelining (Table 9, page 86)come from two other meta-analytic reports; one is by Agent et al. (1996) the other by Gan et al. (2005). The references in Agent et al. are to the Stewardship reports or to older studies. The references in Gan et al. (as best I could ascertain) are to collections of accident reduction factors produced by states, i.e. to meta-analyses that usually do not say how estimates were produced or what their sources are.

In sum, the 'Synthesis" misleads the reader to believe that researchers found edgelining to be good for safety, that evidence shows that retroreflectivity matters because edgelining saves only nighttime accidents, that these results are *"robust*". But this impression is false. From Musick and Basile the 'Synthesis' selected only those results which suited its mission and kept mum about findings which negate it. The 'Stewardship Report', is based on bad data and a demonstrably incorrect assumption. From the valiant effort by Bali et al. Miller manufactures an indefensible

[74]Jackson was the European Highway Marketing Manager for Potters-Ballotini Ltd.

[75] Tamburri et al. say that "There is a remarkable stability present which indicates that in no category did the striping influence accident rates." (Page 67). The newer references in Elvik and Vaa (2004) are about widening edgelines, shoulder strips, detection distance and color coding.

result. What Miller copied from Potters is based on old, methodologically deficient studies, at times from an anecdote or even from no evidence at all.

I limited myself to critically examining what the authors of the 'Synthesis' elected to introduce into evidence. This leads me to conclude that the picture the 'Synthesis' paints is false and that the evidence it musters raises more questions about the role of retroreflectivity than it settles. In short, the evidence in the 'Synthesis' does not support the claim that requiring some minimum retroreflectivity levels will enhance safety.

All this does not mean that edgelining does not save accidents. But impartial reviews of the extant evidence counsel caution. Thus, e.g., based on a review of both empirical and theoretical research findings, Sharfi and Shinar (2014) conclude that visual enhancement systems such as edgelines, raised pavement markers and reflectorized posts may in some circumstances diminish safety.

APPENDIX B. THE 'SYNTHESIS' ON SPEED

To bolster the impression that edgelining must be good for safety the 'Synthesis' aims to show that drivers do not react to edgelining (or to enhanced retroreflectivity) by modifying their speed. To this end, in a separate section (Pages 10 and 11), the 'Synthesis' summarizes the research findings by saying that "…most (studies) show no significant effect in absolute speed difference or, perhaps more importantly, speed variance."

Were this true it would help the FHWA to show that that enhanced pavement marking retroreflectivity is good for safety and do so in two ways. One is by giving credence to the no-effect-on-daytime-accidents assumption on which the findings of the 'Stewardship Report' rests. For, were edgelining found to affect daytime speed then the no-effect-on-accidents assumption would clearly not be plausible. The other way is by forestalling the potentially damaging argument that in some circumstances edgelining (or an increase in retroreflectivity) might make drivers go faster and thereby harm safety.

When reviewing the meta-analysis by van Driel et al. (2004) the 'Synthesis' says that: "The authors came to the conclusion that the net speed effect (of edgelining) was essentially zero."(Page 10). This is not true. On the contrary, van Driel et al. find that edgelining affects both speed and lateral position; it is only that the size of the effect depends on a variety of circumstances. Thus, e.g., van Driel et al. say that "Applying an edgeline to a road without a centerline increases the speed of road users, and replacing a centerline by an edgeline decreases the speed." (Page 671). Van Driel et al. find that the effect on speed may depend on whether it is day or night, whether the segment is straight or curved, on the traffic flow, etc. In one circumstance van Driel et al. report that speed was found to decrease by as much as 5.0 km/h and in another to increase by as much as 10.6 km/h. It is only when one averages over all the circumstances in which van Driel studies the effect of edgelining on speed that "… the mean effects of an edgeline nearly equal 0, both on speed and lateral position." (Page 674).

About another study (Sun and Tekell, 2005) the 'Synthesis' says that: "Conclusively, the researchers found that the addition of an edge line on narrow two-lane highways did not impact vehicle speeds, day or night" This, again, is true only if one speaks about the average over all the circumstances but is untrue for any specific site[76].

In its review of Tsyganov et al. (2006) the 'Synthesis' claims that they found that:"…there were no significant differences in vehicle speeds before and after adding edge lines to the narrow highways … there were no statistical differences in vehicle speeds when considering daytime versus nighttime conditions" and that"… speeds slightly increased in all conditions after edge lines were applied, but the differences were not deemed statistically significant." (Page 11). This too is false.

Based on stationary observations Tsyganov et al. found that on narrow (9') roads, during daylight, speed increased (statistically significantly) by 5.1 mph and during darkness speed decreased(?[77]) by 0.4 mph (Table 3.2, page 28). Based on test drive observations Tsyganov found that "After edge-

[76] Thus, e.g., at Site 1 (tangent, good pavement condition, 5" edge drop, page 85) the average speed dropped from 64.63 mph before edgelining to 64.17 mph after edgelining; at Site 2 (curve, good pavement, outer lane, page 86) the average speed increased from 49.89 mph to 55.51 mph, at Site 3 (tangent, poor pavement, reflectors on centerline, very little traffic, page 87) speed dropped from 41.05 mph to 39.25 mph, …, at Site 9 (S-curve, good pavement, short sight distance, outer lane, advisory speed of 45 mph) the average speed dropped from 39.71 mph to 37.41 mph. Recall also that small changes in average speed are associated large changes in accidents. To illustrate if at site 3 speed decreased by 2% fatalities would be expected to decrease by about 7% and injuries by about 5%.

[77] On page 29 Tsyganov et al. say that after edgelining speed increased by 0.4 mph rather than decreased by 0.4 mph. It is unclear which is correct.

line placement, speed increased for both day and night conditions by approximately 7 percent or 4 mph, on the average, for all drivers and investigated highway classes." (Page 37). In their abstract Tsyganov et al. say that after edgelining there was an "increase speed on average by 5 mph or 9 percent on both straight and curved highway segments".

In sum, while the 'Synthesis' aims to create the impression that edgelining has an insignificant effect on speed the very evidence it invokes shows the opposite[78].

[78]While the 'Synthesis' reviews what Miller (1991) said about the safety effect of edgelining it does not mention what Miller said about the effect of lane markings on speed. Miller assumed that *"The average 56-mph speed on these roads would fall to 54 mph during the peak travel period if the roads were lacking lane lines, edgelines, and centerlines (*Page 42)*".*This, in Miller's equation, is a benefit. It is a peculiar benefit inasmuch as by going faster more drivers exceed the legal speed limit.

APPENDIX C. THE 'SYNTHESIS' COMMENTS ON BAHAR ET AL.

The 'Synthesis' found the evidence that enhanced retroreflectivity is associated with fewer crashes to be sparse and inconclusive. There was, however, one research report (Bahar et al, 2006) with clear results. The basic idea behind the Bahar et al. study was simple. The retroreflectivity of pavement markings fades over time and is periodically restored by repainting. One can therefore check how the accident frequency of roads changes as a function of the decline in retroreflectivity in time. Were one to find that accident frequency increases as retroreflectivity declines that would be a sign that retroreflectivity matters, were one to find that accident frequency drops after repainting that too would be a sign that retroreflectivity matters..

Using data from California, Bahar et al. show that *"the safety difference between high retroreflectivity and low retroreflectivity markings during non-daylight conditions and on non-intersection locations was found to be approximately zero, for all roads that are maintained at the level implemented by California."* They further note that *"Our study provides a level of certainty that builds upon previous research such as Lee et al. (1), Migletz et al. (2), and Cottrell and Hanson (3) which were unable to identify any relationship between retroreflectivity and nighttime crashes*[79]*"* (Page 2) and that *"In summary, this study found that there is no safety benefit of higher retroreflectivity for longitudinal markings on non-intersection locations during non-daylight conditions for roads that are maintained at the level implemented in California's state highways." (Page 4).*

Such findings deny the benefit of what Congress instructed the FHWA to do. The FHWA must have hoped that the 'Synthesis' will find what is wrong with these finding and thereby reduce their credibility. As can be expected, the 'Synthesis' says that the Bahar et al. research has *"significant limitations*"(Page 12) and that "*Combined, these limitations and concerns seriously challenge the quoted concluding remarks …."* (Page 13)[80]

The 'limitations' which the 'Synthesis' finds are five. The first limitation is that Bahar et al. used data from only one state (California). I reviewed several North American studies that attempted to isolate the safety effect of retroreflectivity from that of other circumstances (Hauer, 2017a). All used data from one state only[81]and none, including those by the senior author of the 'Synthesis', mention this to be a limitation. Perhaps they did not see it as a limitation because, while conditions do differ from state to state, it is unlikely that the effect of retroreflectivity on nighttime driving in one state is substantively different from that in another.

The second limitation according to the 'Synthesis' is that Bahar et al. used a data-based regression model to estimate what retroreflectivity prevailed at the time and place when and where an accident occurred while other researchers used measured retroreflectivity levels. This criticism is moot. Retroreflectivity measurements are only seldom available for the location and time proximate to those of an accident. This is why those who used measured retroreflectivity levels had to invent various judgment-based 'imputation rules'. To illustrate, in Michigan nearly all retroreflectivity measurement are taken in September-October, at the end of the remarking season. To get a 'temporarily imputed' value for, say, April, the researchers used the measured value from the previous end of remarking season and applied a fixed monthly degradation value (say, -22 mcd/month for white edgelines). If the result was negative, the retroreflectivity was set to 50

[79] None of these research studies is mentioned in the 'Synthesis'.

[80]Carlson et al. (2013, page 59) repeat these criticisms and amplify by saying that they are sufficiently grave to *"limit (the) acceptability"* of the zero-effect conclusion.

[81] Two studies used data only from Michigan, two used data only from North Carolina, and two used data only from Ohio.

mcd/m^2/lx. .Similarly, if a road segment had no measured retroreflectivity value (covering 40-50% of all cases) an imputed value from within 2 miles was used. It is moot whether this kind of approach yields more appropriate retroreflectivity values than using a regression in which time since repainting, traffic volume and other variables are used to predict what retroreflectivity prevailed on a certain road in a certain month.

The third limitation noted in the 'Synthesis' seems to be a misconception. The 'Synthesis' claims that Bahar et al. study presuppose that retroreflectivity never declines below a level that can adversely impact safety. No such a presupposition is built into the retroreflectivity prediction equations.

The fourth alleged limitation is rather technical. Bahar et al. examined the safety effect of retroreflectivity by grouping roads segments into bins with similar retroreflectivity values trying to ensure that each bin contains enough data for reliable safety effect estimation. The authors of the 'Synthesis' think that the bin boundaries should have been set differently. This limitation is without merit.[82] In any case, since Bahar et al. show that retroreflectivity is not associated with crash frequency in any of the chosen bins, changing their boundaries would not have changed their conclusion.

The fifth limitation which the 'Synthesis' notes is that in California retroreflectivity is only seldom less than 100 mcd/m^2/lx and therefore, if significant safety deterioration occurs only when retroreflectivity is lower than that, the zero-effect conclusion by Bahar et al. would not apply to such low level. This, I think, is a fair observation at least for white pavement markings.[83] Of course, the obverse is also true. Since data with white marking and retroreflectivity above about 100 mcd/m^2/lx and with yellow markings above 80 mcd/m^2/lx were plentiful, the zero-effect conclusion stands in all these bins.

Faced with this evidence, and clutching at straws, the 'Synthesis' speculates that it is perhaps only when retroreflectivity is less than some low value (say, 100 mcd/m^2/lx for white markings), which is where Bahar et al. do not have enough data, that there might be an increase in accident frequency. Such a speculation goes against its own reasoning[84]. Retroreflectivity is known to affect pavement marking visibility measures such as 'detection distance', 'preview time' , 'visibility rating' etc. All such measures are an increasing function of retroreflectivity. The 'Synthesis' asserts (on page 12) that *"Detection distances are thought to be a surrogate for crash data in that longer detection distances have a positive effect on vehicle-control measures and, consequently, crashes. "*If so, how is it that increasing retroreflectivity (and thereby detection distance) above the limiting value does not benefit safety but decreasing it below that limit increases accident frequency?

Bahar et al. asked "… how non-intersection, non-daylight (night, dawn, and dusk) safety is impacted by the change in retroreflectivity of longitudinal pavement markings and markers." (Page 4). In answer they say, correctly in my opinion, that within range of retroreflectivities found in California there is no evidence that more reflectivity is associated with fewer target accidents. None of the limitations noted by the 'Synthesis' casts a reasonable doubt about the validity of their

[82] The authors of the 'Synthesis' think that the bin boundaries ought to be a logarithmic function of retroreflectivity because *"…the performance of retroreflectivity has been repeatedly shown to be best modeled logarithmically."* ('Synthesis', page 13 of 21 and also in Carlson et al. on page 59). It is not clear what is meant by *"performance of retroreflectivity".* Judging by their supporting references and by what is said elsewhere in their paper they probably meant that the relationship between retroreflectivity and 'detection distance' is 'logarithmic'. But why should 'detection distance' govern the choice of bin boundaries? Perhaps lurking behind this argument is the unsupported belief that 'detection distance' and target crash frequency go hand-in-hand.

[83] For yellow markings the lowest retroreflectivity bin is 15-79mcd/m^2/lx. Here too Bahar at al. found no effect on accident frequency.

[84] The 'Synthesis' has a section entitled "Studies of Visibility in Terms of Detection Distances" (Pages 13-14).

findings. The assertion in the ‘Synthesis’ that "Combined, these limitations and concerns seriously challenge the quoted concluding remarks …." is unsupported.