ON THE USE OF REAL-WORLD ACCIDENT DATA FOR ASSESSING THE EFFECTIVENESS OF AUTOMOTIVE SAFETY FEATURES

METHODOLOGY, TIMELINE AND RELIABILITY

Jens-Peter Kreiss
Technische Universität Braunschweig
Germany
Michael Stanzel
Robert Zobel
Volkswagen Group Research
Germany
Paper Number 11-0054

Abstract:
During development of innovative automotive safety features (and therefore well before market introduction) it is common practice for OEMs and their suppliers to do predictive analyses of the anticipated benefit of these systems. It is also common practice that stakeholders do a retrospective analysis once the system in focus is in production. Real-world data is then used to identify the "true" effect of the new system. There are however certain constraints to this approach. The varying degree they are met explain the difficulty to find consistent results and also the time span it takes before such results can be taken with any degree of certainty. It is therefore not surprising that even for systems that are now widely recognized as highly efficient it has taken several years before effectiveness numbers turned out to be stable.

INTRODUCTION

During development of innovative automotive safety features (and therefore well before market introduction) it is common practice for OEMs and their suppliers to do predictive analyses of the anticipated benefit of these systems. These analyses are usually limited to "what-if" simulations which is why they are not affected by take rate / fleet penetration issues. Moreover all parameters in the study can be controlled, i.e. the isolated effect of a single system can be easily shown. On the other hand however assumptions, e.g. about long-term driver behavior adaptation, and other simplifications are usually needed even if their validity can not be guaranteed at this stage.

It is therefore also common practice that stakeholders do a retrospective analysis once the system in focus is in production. Real-world data is then used to identify the "true" effect of the new system. There are however certain constraints to this approach:

- the system in focus must be sufficiently frequent in the real world accident data for its effects to be visible.
- it must be possible to distinguish vehicles with the system from vehicles without the system in the accident data since any misclassification of vehicle equipment or accident situation affects the calculated effectiveness. An estimate of the effectiveness can only be reliable if the nature of misclassifications is non-systematic and its extent is limited.
- the effect of a system that is deployed slowly and has small take rates can be "overwritten" by other systems or concurrent developments.
- confounding factors like belt use must be known (minor changes in the rate of non-use can have dramatic effects on fatality rates).
- case group and control group should differ only in terms of presence of the safety feature in focus. Other parameters like distribution of gender, age, vehicle type, mileage etc. should match. If this is not possible multivariate statistical analysis is needed to identify the respective influence of these parameters.
- results obtained from national or regional studies can not necessarily be applied to predict the effectiveness for regions with different fleet, driver population or infrastructure.
- accident data reflects only accidents that have actually happened, i.e., accidents avoided by a certain technology do no longer appear in the databases. This influences the baseline values.
- over time, drivers may get accustomed to the system and change their driving behavior.
- effectiveness figures published by different authors vary widely in terms of the measure they quantify – skidding, loss of control, single ve-

hicle accidents. This makes results hard to compare.

These constraints and the varying degree they are met explain the difficulty to find consistent results and also the time span it takes before such results can be taken with any degree of certainty - systems like antilock brakes and daytime running lights have been around for decades but are still subject to discussion. It is therefore not surprising that even for systems that are now widely recognized as highly efficient (such as ESC) it has taken several years before effectiveness numbers turned out to be stable.

Electronic stability control (ESC) has been developed to assist the driver of a vehicle in critical loss-of-control situation which may lead to a (in many cases serious) accident. A relevant question of course is whether this wanted behavior of ESC can be confirmed from real world accident data. Moreover a thorough quantification of the effectiveness including confidence limits is necessary. Up to today the corresponding literature on vehicle safety contains quite a number of papers dealing with this question. The assured finding of today is that ESC really constitutes a primary safety system which is of high effectiveness in critical loss-of-control situation. The focus of the present paper is to investigate the history of the corresponding research on the quantification of the effectiveness of ESC with special focus on the question how the confidentiality of the results developed over time.

AVAILABLE STUDIES

Sferco et al. (2001) is an early paper dealing with the potential effectiveness of ESC in Europe. The paper itself states that in 2001 *it was impossible to undertake a fleet study, i.e. a comparison of accident rates between two fleets of similar cars, one equipped with ESP and the other one not equipped, because at that time too few passenger cars on the road were equipped with ESP. Instead the paper presented an estimate of the potential effectiveness of ESP by analyzing a sample of accidents and relying on experts' opinion deciding, case by case, whether ESP would have potentially influenced the process of the accident or not* (cf. Sferco et al. (2001), p.3). The outcome in the paper on the basis of the so-called EACS (European Accident Causation Survey) data was that ESC could have an influence in about 34% of fatal accidents and 18% of injury accidents. But at that time in 2001 a reliably quantification of the possible effectiveness of ESC was not at all possible.

In November 2002 a press release of Mercedes-Benz (also see Unselt et al. (2004)) again indicated that ESC should be regarded as a driver assistance system which may have a significant impact on of loss-of-control accidents. ESC is standard equipment of all Mercedes-Benz passenger cars since summer 1999. So the company compared the two groups of Mercedes passenger cars first licensed in the model year 2000/2001 with those first licensed in 1999/2000.

On the basis of a large random sample of police recorded accident data (including not only injury accidents but property damage accidents as well) for the years 1998 – 2001 from the German Federal Statistics Office it could be observed that among all Mercedes-Benz passenger cars the share of those involved in loss-of-control accidents decreased more rapidly than the share of vehicles involved in loss-of-control accidents among all accident involved passenger cars of other brands. More precisely it was observed that for all brands (excluding Mercedes-Benz) the share of passenger cars involved in loss-of-control accidents decreased from 14.5% in 1999 to 14.3% in 2001. The corresponding share for Mercedes-Benz passenger cars only reduced from 15.0% in 1999 over the same time period to 10.7% in 2001.

It is argued in the press release that the reason for the much faster reduction of the share of loss-of-control accidents in Mercedes-Benz passenger cars compared with passenger cars from other brands is mainly due to the ESC safety system. It is worth mentioning that the press release does not contain a serious statistical quantification for the effectiveness of ESC and that only a reduction of a share was reported upon which in theory of course also could be a consequence of an increasing number of others than loss-of-control accidents. Nevertheless the study strongly indicates that ESC might be an effective system avoiding reasonable parts of loss- of-control accidents. But further and much more detailed investigations would be necessary to obtain reliable and statistically significant quantification of the indicated effectiveness of ESC.

So far all investigations obtained rather crude investigations about the possible effectiveness of ESC. No one of the so far reviewed studies take further variables into account. So it is very clear that most vehicles differ not only according to the equipment with ESC or not, but also typically on other equipment. For example at that time ESC equipped vehicles typically belong to an upper (luxury) segment of the car fleet on the roads. Moreover ESC-equipped passenger cars – especially in the year 2002 and earlier – are much newer than vehicles not equipped. Questions of the possible effects of vehicle mass, road conditions, gender and age of the driver and more general driving behavior in different categories of vehicles and so on, have so far not been taken into account.

In the year 2003 a first much more quantitative study of the effectiveness of ESC was presented on the ESV conference in 2003 by Tingvall et al. (2003). This study uses police recorded accident data from Sweden with at least one injured person from the years 2000 until 2002. Only vehicles from the model years 1998 – 2003 were included in the study. The authors build two groups of various case (ESC equipped) as well as control cars (not equipped with ESC) where the controls were selected to be as close as possible to a case vehicle. In total 442 case cars and 1967 control cars are considered in this study. The main idea of the authors is to compare the share of ESC-equipped vehicles in the two accident groups of rear-end accidents and all accidents except rear-end with each other. The reason for this is, that it is assumed that the equipment with the safety function ESC more or less has no potential to affect rear-end accident situations.

Under this assumption the share of ESC-equipped passenger cars in the group of rear-end accidental situations could be regarded as a reliable estimate of the share of ESC-equipped passenger cars on the roads. And of course this share is the quantity to which we can compare the share of ESC-equipped passenger cars among other accidental situations in order to see whether there is a reduction or not. A reduction of the share of ESC-equipped vehicles in a specific accident situation (other than rear-end) will be interpreted as an effectiveness of ESC with respect to the specific accidental situation considered.

To understand better this quite important approach let us consider a simple example, taken from Kreiss et al. (2005) and coming from real accident data. From a large sample of about 690'000 police recorded passenger accidents of the German Federal Statistical Office for the five years period 1988 – 2002 we extract only fatal accidents in which passenger cars are involved for which we most likely know whether these vehicles have been equipped with ESC or not. Moreover we only include involved passenger cars into the study if we most likely know whether ESC has some potential effect on the accident outcome of this specific vehicle (accident sensitive to ESC) or if ESC definitely has no effect on the accident outcome (accident not sensitive to ESC). Doing so we end up from our huge accident data base with a sample of n=432 passenger cars involved in fatal accidents. The results can clearly be arranged as shown in Table 1..

Exactly as described in Tingvall et al. (2003) we use the equipment ratio of 54/68=79.4% for the vehicles involved in accidents not sensitive to ESC to estimate the ESC equipment ratio for the vehicles on the road. It is stressed that the quantity 79.4% only is an estimator of the unknown equipment ratio on the market and that this estimator may suffer from systematic (e.g., vehicle selection) as well as completely unsystematic (probabilistic) fluctuations. Since the equipment ratio of 82/228 = 36.0% for the vehicles involved in sensitive accidental situations is much lower we have an indication of a relevant effect of ESC.

Table 1.
Cross-tabulation in Kreiss et al (2005)

Fatal accidents	Vehicle not equipped with ESC	Vehicle equipped with ESC	Total
Accident not sensitive to ESC	68	54	122
Accident sensitive to ESC	228	82	310
Total	296	136	432

For a precise and correct quantification of the effectiveness of ESC note that according to our assumptions ESC only can have an effect on the category of sensitive accidents in ESC-equipped passenger cars. That is only the number 82 in the above Table 1 might be influenced by ESC. Under the assumption that ESC completely has no effect we would expect the same equipment ratio of 79.4% for the passenger cars involved in accidents sensitive to ESC. If this would have been the case we would have expected a number x instead of 82 accidents in Table 1 such that the fictive equipment ratio x/228 equals 79.4%. This leads to x=181. Thus ESC was able to reduce the fictive number of accidents sensitive to ESC from 181 to 82 which constitutes an impressive reduction rate of 54.7% for ESC in the category of accidents sensitive to ESC. Rewriting the reduction rate as

$$\begin{aligned}\text{Effectiveness} &= 1 - \text{OR}\\ &= 1 - (82*68) / (54*228)\\ &= 1 - 0.453 = 0.547 = 54.7\%,\end{aligned}$$

where

$$\text{OR} = (82*68) / (54*228)$$

is the so-called Odds-Ratio of Table 1.

This exactly describes what Tingvall et al. (2003) did. They considered as the category of accidents not sensitive to ESC rear-end accidents on dry roads. This means that the results assume that rear-end accidents on dry roads are not expected to be influences by ESC. This assumption is justified as long as direct influence is

considered. However ESC is usually combined with some kind of brake assist (BAS). Such a system may have a influence on rear-end accidental situations, so there might be an indirect effect of ESC via the link to BAS. Fortunately it can be seen that such a possible indirect effect typically leads to an underestimation of the effectiveness of ESC. (cf. Kreiss et al. (2005)).

As accidents sensitive to ESC the Swedish study reported upon in Tingvall et al. (2003) selected all accidents except read-end on dry roads. This means that this study, in contrast to our example from above, considers the effectiveness of ESC to all accidents except rear-end on dry roads.

Since the paper of Tingvall et al. (2003) does not give explicit tables like Table 1 but results of the effectiveness only we could not have used their data for our example above.

The confidence interval [1.1% , 43.1%] for the overall effectiveness of ESC on all but rear-end accidents given in the paper of Tingvall et al. (2003) is quite large. In other words, with 95% confidence we can only state a 1.1% effectiveness of ESC. Concerning the effectiveness on all but rear-end accidents on dry roads the paper states a 95% confidence interval of [-19.0% , 37.6%], which does not state a significant effect of ESC on dry roads at all! On wet roads the study obtained a 95% confidence interval of [7.8%, 55.2%] for the effectiveness of ESC.

The authors stress that because of the different weather conditions it is impossible to generalize the results of their specific study to all other parts of the world. Moreover it is remarkable that the Swedish study cannot detect a significant effect of ESC on dry roads. Finally the Swedish study does take into account varying road conditions.

In the paper Aga and Okada (2003) the authors report on a study of the ESC effectiveness in Japan. This study not only took accident material (provided by ITARDA) into account. Instead the rate or risk of suffering an accident when driving on the roads is considered and the study gives estimates for this risk of passenger cars equipped and not equipped with ESC by computing the numbers of accidents per 10'000 registered vehicles per year. It is obtained that vehicles equipped with ESC have a of about 35% lower risk of suffering a single car accident. The investigation and computation of this risk rates was done in such a way that vehicles registered for the first time in 1994 were investigated according to their accident behavior for the period 1994 – 1998, vehicles registered for the first time in 1995 have been under accident inspection from 1995 until 1999 and so on. Since ESC have been introduced somewhere in between the study we have to face the fact that in the study older vehicles not equipped with ESC and more recent and modern vehicles equipped with ESC have been considered. And of course and also as already mentioned these passenger vehicles of varying year of manufacture differ according to their safety equipment in more than ESC. Even the mileage per year and this means the time under risk on the roads may differ for the vehicles and would lead to a biased estimator of the reduction of the accident rate! However the paper of Aga and Okada (2003) presented a different and relevant approach for investigating the possible effectiveness of ESC.

In a further paper, Page and Cuny (2004) report on a study similar to the approach of Tingvall et al. (2003) for French cars. In this study the authors compared the Renault Laguna 1 (which was not equipped with ESC) with the newer vehicles Laguna 2 – equipped with ESC. The selected sample consists of 1356 cars involved in injury accidents in 2000 – 2003 in France. In contrast to Tingvall et al. (2003) the authors have thoroughly selected a variety of accidental situations relevant to ESC and/or braking. Even more the authors considered several different scenarios of driver age, vehicle age or year of accident. As described above in detail the authors give 1 – Odds-Ratios as estimates of the effectiveness of ESC in accidents sensitive to this system. Without taking into account any confounding factors (crude approach) Page and Cuny (2004) for example obtain the following 2x2 contingency table.

Table 2.
Cross-tabulation in Page and Cuny (2004)

	ESC equipped	**No ESC**	**Total**
ESC sensitive accident	22	177	199
Non ESC sensitive accident	71	318	389
Total	93	495	588

which leads to a (crude) Odds-Ratio of OR = (22*318) / (71*177) = 0.56 indicating that 1 – OR = 44% of all ESC sensitive accident situations could be avoided by the electronic stability program. A corresponding 95% confidence interval for the (crude) Odds-Ratio was given in the paper as [0.46 , 1.29] and meaning that the value OR = 1 could not be excluded on this significance level. This means that we unfortunately do not have a significant effectiveness of ESC on French roads from this study. Of course this is due to the fact that the sample size in this study as well as in all other studies so

far is limited, even so a period of several years of police recorded accidents have been taken into account. The reduction of the accident sample because of the necessity of having knowledge of the equipment with the safety function as well as the selection of sensitive and not sensitive accidents dramatically leads to a reduction of the sample size. This holds true for all studies even in large countries and is not a drawback of the study of Page and Cuny (2004). A very relevant result from Page and Cuny (2004) is the fact that here for the first time we obtain indications that the effectiveness of ESC varies with confounding variables like driver age, vehicle age and year of the accident. Again these results are not statistically significant. A certain drawback of the study of course is that the comparison of the Laguna I and its successor Laguna II leads to the problem that these two passenger cars differ in much more than the equipment with ESC, only.

It shows that with an error probability of 5% we can expect an effectiveness of the new Laguna 2 in comparison with the old Laguna 1 according to ESC-sensitive accidents of at least 7 %. Of course this reduction rate of 7% is only valid for the group of ESC sensitive accident constellation and by no means for all accidents as in the study of Tingvall et al. (2003).

Let us see how sensitive the results reported upon in Table 2 are. Sources for uncertainty are

- precision of the ESC-equipment ratio in the category of non ESC sensitive accidents as a surrogate fort he equipment rate on the roads
- Misclassification of vehicles according to equipment
- Misclassification of accidents according to type
- Under- or over-reporting of certain accident types
- Influence of additional factors like age or gender of driver, driving behavior depending on vehicle category
- Comparability of vehicles equipped and not equipped with ESC

To investigate possible effects of these facts assume that the equipment ratio of 71/318=22.3% in the category of non ESC sensitive accidents systematically under- or overestimates the equipment ratio on the roads by about 10%, that is hat the equipment ratio on the roads may be either 24.8% or 20.1%. This would immediately lead to a quite substantial uncertainty about the effectiveness of ESC within the range of (38%, 50%).

As a further example let us assume that about 10% of the reduction of ESC-sensitive accidents in Table 2 are completely explained by further safety features (other than ESC) and that only the ESC equipped vehicles additionally are equipped with this functions. This would lead to the following modification of Table 2 concerning the pure effectiveness of ESC.

Table 3.
Cross-tabulation by Page and Cuny (2004), modified to account for non-ESC effects

	ESC equipped	No ESC	Total
ESC sensitive accident	24	177	199
Non ESC sensitive accident	71	318	389
Total	93	495	588

The same calculation of above would lead to a decrease of the computed effectiveness of ESC to 38%.

Misclassification of vehicle equipment and/or accidental situations of course has an effect on the quantification of the effectiveness of ESC. From Kreiss et al. (2005) it is seen that a completely random misclassification of both mentioned types has the effect that we underestimate the effect of ESC. So this in a sense is not that problematic. Much more delicate would be a situation in which the used accident data base contains some systematic errors or reporting rates of only one type of accident or only one group of ESC- or non ESC equipped vehicles. This would lead to a systematic variation concerning the quantification of the effectiveness of ESC.

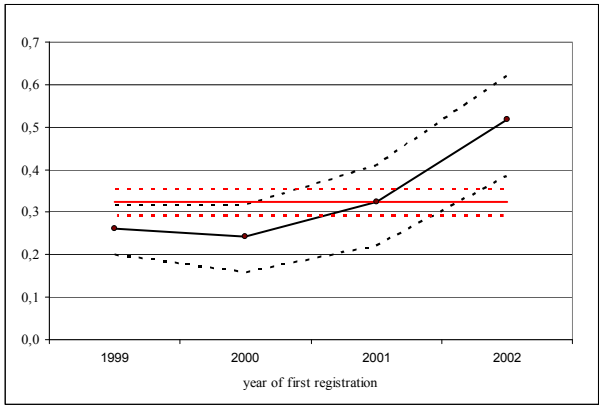

Figure 1. Effectiveness of ESC for different years of first registration (1999-2002) (black) with 95% confidence limits (dotted) and overall effectiveness including 95% confidence limits (red)

Further difficult to detect sources of uncertainty in the quantification of the effectiveness of ESC may be

hidden in additional and relevant variables. Some examples are discussed in Kreiss et al (2005). From the already mentioned rather large sample of German accidents for the period 1998-2002 this paper reports that the effectiveness of ESC varies according to a variety of additional variables. For examples the following result is obtained concerning different years of first registration

Figure 1 indicates that either the functionality of ESC systems in vehicles improved over the years or that additional safety equipment in more recent cars has some effects on loss-of-control accidents or even both. Another possible explanation is that the underlying accident data most probably is sampled during the same period, i.e. the 1999 cars may simply be older (and therefore driven by a different driver population) than the 2003 ones.

Concerning the gender of the driver Kreiss et al. (2005) obtain a significantly better effectiveness of ESC in women-driven vehicles, cf. Figure 2.

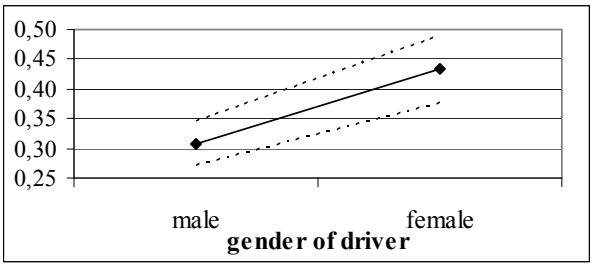

Figure 2. Effectiveness of ESC separately for gender of driver including 95% confidence limits

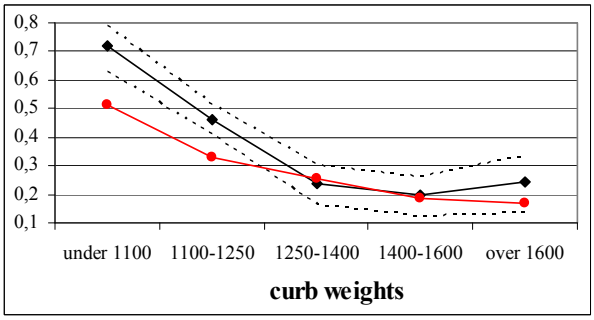

Figure 3. Effectiveness of ESC separately for different curb-weights (in kg) including 95% confidence limits (black solid and dashed lines) together with the percentage of female drivers in the respective curb-weight category (red line)

Since it can be verified at least from accident data that women on average drive smaller sized vehicles than men the effect seen in Figure 2 may be to some extend or even completely be explained by a possible variation of the effectiveness of ESC according to vehicle size. To this end consider Figure 3 presenting the variation of the effectiveness of ESC in comparison to vehicles size and also the percentage of women-driven vehicles (involved in accidents) according to the different vehicle category.

Finally Kreiss et al. (2005) obtain from the German accident data base that ESC is more effective on dry than on wet or icy roads which is in contrast to the Swedish results of Tingvall et al. (2003) and underpins the remark of the Tingvall paper that results from one country can not easily be extended to other countries!

Weekes et al. (2009) claimed a higher effectiveness rate of ESC for young drivers in the UK. More precisely the study states that the overall effectiveness of ESC for young drivers of about 14% is around double a previously published overall effectiveness of 7% for all ages and all injury severities. Since the paper also states that young drivers commonly drive small used cars with ESC rarely fitted the question arises how far the reported effectiveness is related for example to the smaller vehicle, since ESC if equipped in smaller and lighter vehicles might have a dramatically higher effectiveness (cf. Section 2 and Figure 3).

ESC appeared on the U.S. market a few years later compared to Europe. Farmer (2004) compared on police-reported crashes for seven states from the years 2001 and 2002 crash rates per registration for selected group of vehicles. The main focus was on vehicle models which changed from no ESC or optional ESC to standard ESC in consecutive model years. Concerning the overall crash involvements and the injury crash involvements the papers describes slight effects (which not in all cases are statistically significant), only. Concerning fatal crash involvements the observed effects are larger. A closer look on the reported results shows that the picture is indifferent over the considered vehicle models. For some ESC equipped vehicle models even higher numbers of fatal crashes than expected where observed.

Another inconsistency, likely based on little data, is that the observed effectiveness of ESC was larger in the group of vehicles which changes from optional to standard ESC equipment compared to the group of vehicles for which the equipment changes from no ESC to standard ESC. However the observed effects of for example 41% lower than expected number of single vehicle crashes is quite similar to reduction rates reported for Japan or Europe. Concerning the overall multiple vehicle crash-rates Farmer (2004) reported a little, if any, effect of ESC, which is in contrast to the

studies of Aga and Okada (2003) and Tingvall et al. (2004).

Dang (2004) on the basis of single vehicle crash data from the years 1997-2002 from five US states in her evaluation note reduction rates of about 35% for single vehicle crashes in ESC equipped passenger cars. A 30% reduction is reported for fatal single vehicle crashes (also for passenger cars). For SUV's the obtained reductions rates are higher. The effectiveness of ESC is computed by using multiple-vehicle crashes as control group. Since the belt usage rate among passengers in ESC equipped SUV's (58%) was higher than among passengers in non ESC equipped SUV's (49%) and the small samples do not lead to reliable results for SUV's.

The paper of Bahouth (2005) reported on six US state files of about 14,000 police-reported crashes for the years 1998-2002 and 11.2% reduction in multiple-vehicle frontal crashes as well as a 52.6% reduction in single-vehicles crash rates for ESC-equipped Toyota passenger vehicle models and SUV's in the US. Following the methodology of Tingvall et al. (2003) the study of Bahouth (2005) compares rear impact crashes with multi-vehicle frontal crashes and singe-vehicle crashes. The obtained results are similar to the findings of Farmer (2004).

AVAILABILITY OF REAL WORLD DATA

1.1 National statistics

All major developed countries publish their annual road accident statistics, usually based on police accident reports. Although concerns about underreporting have been voiced (and corroborated by hospital and insurance data,[]) the national databases are considered the most comprehensive in terms of case numbers. On the other hand however, they allow only limited or no access to disaggregate data and/or omit detailed injury, technical or reconstruction information. Hence, scientific in-depth databases have been established to fill this gap.

1.2 Scientific in-depth databases

These databases cover only a small fraction of all accidents but provide detailed information on the vehicles and their safety features, medical data as well as reconstruction results for the collision and, in some cases, the pre-crash phase. If available, the latter also allow for an analysis of accident causation. There are different sampling philosophies (random, stratified…,) that may or may not allow a projection to national statistics.

GIDAS, the German In-Depth Accident Database is an example of a scientific database. It was launched in 1999 and has BAST (the Federal Highway Research Institute) and FAT (an industry consortium) as sponsors. Two academic institutes, Hanover Medical School and the Technical University of Dresden act as contractors. Each contractor collects about 1000 cases per year, following a common methodology and storing them in a common database. The entry criteria,

- road accident
- involving personal injury (of any severity)
- within defined geographical regions
- while the analysis team is on call

should make sure that sampling is random and hence the sample is representative. This is true for most key variables except accident severity – the more severe an accident is the more likely the GIDAS team will be notified by the police. This causes a certain bias towards severe and fatal cases; however this bias can be corrected by means of weighting factors if necessary.

1.3 OEM proprietary databases

Since scientific databases are often designed to be representative of national statistics the vehicles sampled in them also represent a cross-section of makes, models and model years. Manufacturers trying to establish the real-world safety performance of a new model or feature will however look primarily at accidents involving new (and, if possible well-equipped) vehicles from their own model range. This is why some OEMs have set up their own accident investigation teams, often feeding their information into databases even more detailed than the scientific databases discussed previously. Due to their (intentional) bias in terms of vehicle selection and some other (unintentional but sometimes inevitable) bias in terms of geography, accident severity, etc… projection from these data to national statistics is problematic.

TIME LAG BETWEEN TAKE RATE AND FLEET PENETRATION

Advanced safety features are usually introduced following a top-down approach, i.e. at first they are offered as an optional extra for luxury executive vehicles, then gradually becoming available in more and more family vehicles until, in many cases, the feature becomes standard equipment even in compact and economy cars. Furthermore, advanced features usually cannot be retrofitted, i.e. their fleet penetration can grow only as fast as old vehicles are replaced by new ones.

Both of these effects cause a substantial time lag between the market introduction of a new technology and the time it can be found in significant numbers in the fleet. The following example illustrates this. For simplicity it was assumed that both the overall number of

vehicles in the fleet and the replacement rate do not change over time. This leads to an age distribution like the one in the following graph (slightly idealized from 2009 German registration data):

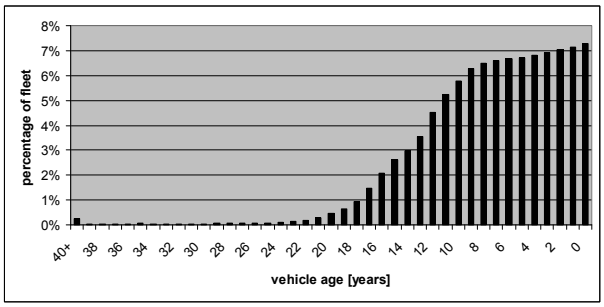

Figure 4. Sample age distribution of a given fleet of cars

Let us further assume that a new promising in-vehicle technology will be phased in according to the following scenario. Please note that a time span of 10 years between market introduction and 100% equipment rate is relatively short by comparison - most systems have taken a longer time, some have never become standard.

1year after market introduction	5% take rate
2 years after market introduction	10% take rate
3 years after market introduction	20% take rate
4 years after market introduction	35% take rate
5 years after market introduction	50% take rate
6 years after market introduction	65% take rate
7 years after market introduction	80% take rate
8 years after market introduction	90% take rate
9 years after market introduction	95% take rate
10 years after market introduction	100% take rate, i.e., standard

After the first year approximately 0.4% of all vehicles in the fleet will be equipped:

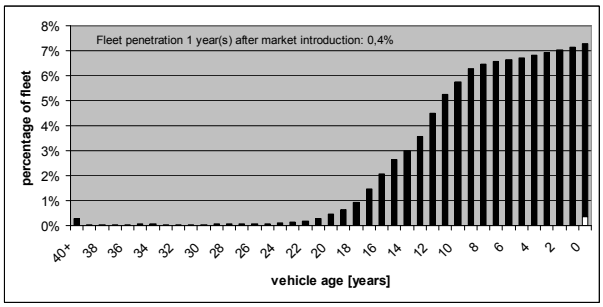

Figure 5. Age distribution and system fleet penetration 1 year after market introduction

Another three years later the fleet penetration will still be as low as 5%:

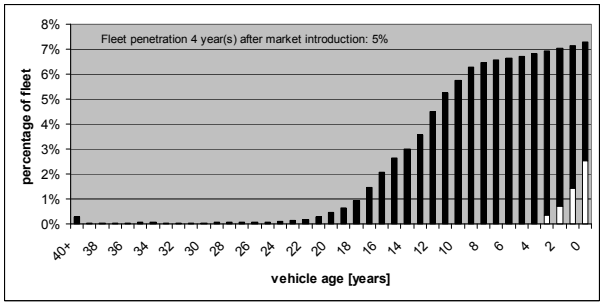

Figure 6. Age distribution and system fleet penetration 4 years after market introduction

Even ten years after introduction, i.e., when we assumed the system to have become standard on every new vehicle there will be many more vehicles without the system than with the system on the road:

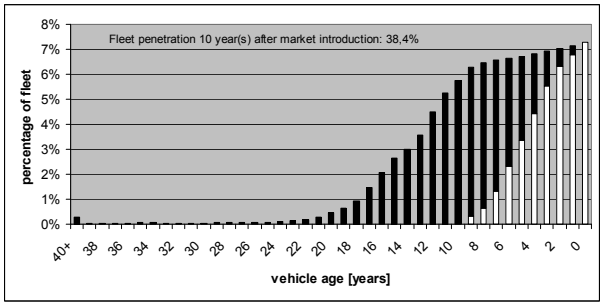

Figure 7. Age distribution and system fleet penetration 10 years after market introduction

This illustrates the difficulty of an early assessment of a newly introduced system, a problem that is often made even worse when the first buyers ("early adopters") of a new technology are not a representative cross-section of the overall population of customers.

IDENTIFICATION OF SAFETY FEATURES IN THE ACCIDENT DATABASES

For a retrospective analysis of a given system's safety benefit it must be possible to identify the presence or absence of this feature in the accident database. Typically, police data do not contain any such information. Given the multitude of makes and models in today's fleet officers on the scene can not tell with certainty which systems have been present. Arguably this gap could be filled with the help of the respective OEM who might (even years after the accident) look up this information in its production history data. As many systems in question are optional their presence needs to be established on a per-vehicle (i,e., VIN) basis. Different OEMs however have different systems to record

their production history which makes this approach tedious. Moreover, forwarding a VIN list to an OEM raises privacy issues, so this approach is not even legally possible in many jurisdictions.

Generally speaking, scientific databases are facing the same problems. Given the lower case numbers however it is often possible for the investigators to gather the required information from inspecting the vehicle on the scene. In this case however only systems will be identified whose fleet penetration is high enough to justify the effort of introducing the respective variable in the database. This adds to the delay described in the previous section.

CORRELATION VS. CAUSATION

One of the most common errors in statistical studies is the confusion between correlation and causation. Something (e.g. a driver action or a safety system) can cause another action or occurrence, which means that the primary action or occurrence really implies or at least has some influence on another action or occurrence. As an example, a damaged breaking system in a vehicle is a possible cause for a rear-end accident in nose-to-tail traffic. And if one action or occurrence causes another, then they are most certainly correlated. Correlation only means that there is some tendency of two or more actions or occurrences to coincide in a (typically complex and random influenced) situation. But just because events occur together this does not mean that one causes the other, even if it may make some sense. If one reduces correlation to the classical statistical correlation coefficient, then correlation even measures linear dependence, only. And it is rather easy to think of variables, which are closely (and even completely) dependent – for example in a quadratic or more complicated nonlinear way – and having a correlation coefficient of exactly zero!

An extreme difficulty in many statistical methods is that they are designed to detect or to test correlation of two or more events only! To obtain evidence that observed correlation comes from causality of two o more factors one has to rule out the possibility that the observed factors are caused by one or more further and typically not observable factors. Strictly and theoretically speaking this really hardly can be done. In some situation causation may be deduced from common or specialists sense, but there are of course many cases in which the existence of causality of events is not so clear. Studies in which only two or a very restricted number of variables are considered are to a large extend not able to give reasonable evidence in direction of causality.

The applied statistical literature recommends so-called controlled studies in order to give evidence that observed correlation is connected to causation. In a controlled study, two or more groups of observational objects (e.g. vehicles with drivers on the road) are created, which in almost every way are comparable (same age and gender of driver, same vehicle and same driving and road conditions and so on) except the one (e.g. a specific safety function) one is interested in. Exactly this is done in serious clinical trials developed in order to detect causation between for example lung cancer and smoking.

For the evaluation of the influence of a specific safety function or a an complete safety equipment in vehicles on the road and based on real-world accident data the statistical tool of a controlled study by far is not an option. The selection of accidences, which to a suitable amount coincide in respect to driving situation, weather conditions, driver and vehicle, in many if not all cases is impossible because of a limited number of accidents available and more seriously because of the lack of information. As a example note that the risk of being rather severely or even fatally injured cannot be reliably computed from an accident database in which the information whether the driver was belted or not is not available.

Because of this one has to be really aware of falsely jumping to causal conclusions too early. This seems to be extremely the case when there is some public suspicion about the reasoning for some observed effects.

As an example let us consider once again the situation reported upon in Figure 2 (Section 2). There we find out a statistically significant correlation between the gender of the driver of a vehicle and the effectiveness of the ESC system in loss-of-control accidents. This result does not mean that there necessarily is some causation between the two factors, more exactly that an ESC system works better in a female driven vehicle than in a vehicle driven by men. Of course there may be causality, but the slightly deeper investigation, which includes the size of the vehicle, shows that gender of driver and the size of the vehicle are correlated as well (cf. Figure 3). So it is quite possible that the true causality is between the size of the vehicle and on one hand the gender of the vehicle (women tend to driver smaller cars) and on the other hand between ESC and size of the vehicle (the safety increment by ESC in smaller vehicles is larger than in larger cars typically having more primary and secondary safety equipment on board than smaller cars). This really may be the case and really may lead to the observed pseudo-correlation between gender of the driver and effectiveness of ESC in loss-of-control accidents. Unfortunately, it may be the case that none of the above is true and the true reason for all observations is much more complex. A further more

complex and still plausible explanation for the observed phenomenon may be that women indeed tend to drive smaller vehicles but also tend to drive in different situations and moreover possess a different driving attitude than men and that (young?) men especially in small cars tend to overestimate their and the vehicles possibilities and are much more exposed to loss-of-control accidents.

All this demonstrates how complex reality might be and possible is. Thus a reliable conclusion from a single accident investigation is difficult if not impossible to obtain. It is the variety of accident-based investigations leading to comparable and therefore reproducible results that forms a convincing picture of a causal dependence of the ability specific safety function to partly avoid specific types of accidents.

In case a statistical database analysis shows a positive correlation between two variables a and b it takes engineering judgement (or at least common sense) to determine whether a causes b, b causes a or both a and b must be attributed to a third effect, cf. Zobel (2007)

DRIVER EFFECTS VS VEHICLE EFFECTS

To demonstrate that driver effects not taken into account may lead to a substantial bias in statistical conclusions let us consider a thought experiment.

Assume that we have n=1.000.000 vehicles on the road and that 30% of all vehicles are equipped with a specific safety function of interest. Let us think for simplicity of just one driver related variable (Driver) which can attain two value, 0 and 1, say, only. An example is gender of the driver. We restrict our investigation to only to accident scenarios of two types, namely one sensitive to the safety function of interest (Sensitive Accident) and the other one more or less neutral (Neutral Accident) accidental situation. The assumed model to generate accident data is a simple logistic regression model of the form

$$P(\text{Sensitive Accident} \mid \text{Safety Function} = r, \text{Driver} = x) = \exp(\beta_0+\beta_1 r+\beta_2 x)/ [1+ \exp(\beta_0+\beta_1 r+\beta_2 x)]$$

for all r,x =0,1 and β_0=-5, β_1=-0.35 and β_2=0.50. This means that we have a positive effectiveness of the safety function as well as of Driver = 0. More precisely the effectiveness of the safety function on accidents sensitive to it reads as follows:

Effectiveness = 1- exp(-0.35) = 0.295 = 29.5%.

Let us further assume that for about 80% of the vehicles Driver =0 is true and that we have the following distribution of the variables Driver and Safety Function

Table 4.
Driver distribution in equipped and non equipped vehicles

	gender		
safety Function	**male**	**female**	**total**
not present	600,000	100,000	700,000
present	200,000	100,000	300,000
total	800,000	200,000	1,000,000

Table 4 reflects that 30% of the vehicles are equipped with the safety function but that this rate varies according to Driver gender equal to “male” or “female”.

According to our assumption on the probability for an accident we obtain by Monte Carlo simulation the following tables of accidents.

Table 5.
Simulated number of accidents sensitive to the safety function

	gender		
safety Function	**male**	**female**	**total**
not present	4,009	1,097	5,106
present	951	779	1,730
total	4,960	1,876	6,836

Table 6.
Simulated number of accidents neutral to the safety function

	gender		
safety Function	**male**	**female**	**total**
not present	4,050	2,100	6,150
present	1,350	2,100	3,450
total	5,400	4,200	9,600

The values in Table 6 accurately resemble the equipment rate within the two driver categories (compare the respective columns in Tables 4 and 6). But, the probability of suffering an neutral accident varies within the two driver groups.

Using the standard SPSS-routine *logistic regression* the following estimators are derived:

β_0 = -0.010, β_1 = -0.341 and β_2 = -0.640.

It can be seen that only the estimator for β_1 and therefore fort he effectiveness of the safety function satisfactorily works, while the estimator for β_2 completely is misleading.

Now let us see what happens if we apply the logistic regression routine without taking the two different driver categories into account, thus what happens if we do not take gender into account. Doing so we end up with the following 2x2 contingency table of accidents

Table 7.
Simulated numbers of accidents

	accident type	
safety Function	**neutral**	**sensitive**
not present	6,150	5,106
present	3,450	1,730

The estimator for the effectiveness of the safety function in this simplified situation without any confounding variables is rather easy and reads as follows

Effectiveness = 1 – 6,150*1,730/(3,450*6,106) = 39.6%

and therefore overestimates the effectiveness by about 34% when compared to the "true" effectiveness of 29.5% (see above).

This simple example demonstrates that it is rather essential to include confounding variables when they may have a non-negligible influence.

Summarizing one can say, that the effectiveness of a specific safety function reliably can be estimated on the basis of real world accident data only if relevant confounders are included in the investigation. Falsely ignored confounders may lead to substantial errors in estimating the effectiveness of a safety function even in very simple examples.

SYSTEM A VS SYSTEM B

As already mentioned at several places in this paper it is most wanted to obtain from real-world accident data reliable and significant results on the causal effectiveness of some safety functions or some safety equipments. To do so and especially to obtain enough reasoning for causality it is necessary to include so-called explaining variables in the investigations in order to be able to (at least partly) control for the influence of these covariables. Age and gender of the driver, weather and road conditions, seat belt usage and vehicle age, mileage and vehicle equipment may serve as examples of covariables, which should be included. So far the theory. In real data applications this to a considerable extend is not possible because of the lack and reliability of information in accident databases.

But even if we assume that we have all these information at hand then in a lot of investigations the total number of accidents tends to be way to small to carry through a very detailed statistical analysis. Of course there exists statistical models, which allow for the inclusion of a lot of variables but most often these models are of so-called parametric nature, like the logistic regression model is. These models and especially the logistic regression model are rather handy to apply to multivariate observed data and the results typically look quite nice and easy to interpret so that one might be tempted to apply these models without further thinking. But a great disadvantage of parametric models is hat they by their definition make very strong assumptions on the underlying dependence structure of the various variables. Of course, when dealing with data a linear dependence is the easiest to assume and of course the obtained results look nice because they are linear. But the assumed model does not allow for anything else than linearity! The same objection is true for the logistic regression model, which assume after a certain logarithmic transformation nothing else but a linear multivariate model. At least this linearity assumption has to be checked on the data before the logistic regression model is applied and conclusions are drawn.

CONCLUSIONS

For a number of years after the introduction of ESC researchers from different regions of the world have published retrospective analyses of this system's safety benefit. Using different data sources, mathematical approaches and effectiveness metrics the overall results were scattered and, in some respects, contradictive. Some studies found that – with the data available at the time – they could not rule out a *negative* effect on safety at the 95% confidence level. This initial uncertainty is caused by various factors, many of them – by nature - beyond the control of the respective authors. Any early study of a new technology's safety effect should therefore be taken with care.

REFERENCES

[1] Aga, M., and A. Okada (2003): "Analysis of Vehicle Stability Control (VSC)'s Effectiveness From Accident Data. 18th-ESV-Conference, Nagoya (Japan).

[2] Agresti, A. (1996). "An Introduction to Categorical Data Analysis". Wiley Series in probability and Statistics (New York).

[3] Bahouth. G. (2005). Real World Crash Evaluation of Vehicle Stability Control (VSC) Technology. 49th Annual Proceedings of the Association for the Advancement of Automotice Medicine.

[4] Becker H., et al. (2004). "Methods for the Evaluation of Primary Safety Measures by Means of Accident Research". Paper F2004V039, FISITA Conference, Barcelona (Spain).

[5] Dang, J. N. (2004). "Preliminary Results Analyzing the Effectiveness of Electronic Stability Control (ESC) Systems". U.S. Department of Transportation – National Highway Traffic Safety Administration. Evaluation Note.

[6] Derriks, H. M. and Peter M. Mak (2007): "Underreporting of road traffic casualties". IRTAD Special report

[7] Farmer, C.M. (2004). Effect of Electronic Stability Control on Automobile Crash Risk. Traffic Injury Prevention 5, 317 – 325.

[8] Hautzinger H. (2003). "Measuring the Effect of Primary Safety Devices on Accident Involvement Risk of Passenger Cars – Some Methodological Considerations". Working paper of the SARAC II project, Heilbronn, IVT.

[9] Hautzinger, H. et al. (1993). "Dunkelziffer bei Unfällen mit Personenschaden" [Number of Undetected Casualty Accidents]. Bericht M 13 der Bundesanstalt für Straßenwesen (BASt)

[10] Krafft, M. et al. (2009). "From 15% to 90% ESC Penetration in new Cars in 48 Month – The Swedish Experience". ESV paper No. 09-421. 21st ESV Conference, Stuttgart.

[11] Kreiss, J.-P. et al. (2005). "The Effectiveness of Primary Safety Features in Passenger Cars in Germany". ESV paper No. 05-145. 19th ESV Conference, Washington D.C.

[12] Langwieder K. et al. (2004). "International Field Experience with Electronic Stability Program (ESP) in cars". FISITA-Conference, Barcelona (Spain).

[13] Martin J.-L. et al. (2003). "Estimating Relative Driver Fatality and Injury Risk According to Some Characteristics of Cars and Drivers Using Matched-Pair Multivariate Analysis". ESV-paper No. 364. 18th ESV Conference, Nagoya (Japan).

[14] Martin J.-L., Derrien Y., Laumon B. (2003). Estimating Relative Driver Fatality and Injury Risk According to Some Characteristics of Cars and Drivers Using Matched-Pair Multivariate Analysis. ESV-paper No. 364. 18th ESV-Conference, Nagoya (Japan).

[15] Mercedes Press Release November 2002

[16] Metzner, G. et al. (1992): "Dunkelziffer der Verkehrsunfalltotenstatistik", Internationales Verkehrswesen 44 (1992), p. 25

[17] Otto, S. (2004). "Quantifizierung des Einflusses aktiver Sicherheitssysteme auf die Unfallwahrscheinlichkeit und Identifikation von sicherheitsrelevanten Attributen basierend auf Realunfalldaten". Diploma Thesis, Universität Dortmund.

[18] Page Y. and S. Cuny (2004). "Is ESP effective on French Roads". 1st International ESAR, Hanover (Germany).

[19] Sferco R. et al. (2001). "Potential Effectiveness of Electronic Stability Programs (ESP) – What European Field Studies Tell us". ESV paper 2001-S2-O-327. 17th ESV Conference, Amsterdam (The Netherlands).

[20] Tingvall, C. et al (2003). "The Effectiveness of ESP (Electronic Stability Programme) in Reducing Real Life Accidents". ESV paper 261. 18th ESV Conference, Nagoya (Japan).

[21] Unselt, T. et al. (2004). "Avoidance of Loss-Of-Control Accidents Through the Benefits of ESP". Paper F2004V295, FISITA Conference, Barcelona (Spain).

[22] Weekes, A. et al (2009). "ESC Standard Fitment and Failure to Protect Young Drivers". Paper No. 09-278. 21st ESV Conference, Stuttgart (Deutschland).

[23] Zobel, R. et al. (2007): "What accident analysis tells about safety evaluations of passenger cars", Paper Number 07-0074, 19th ESV conference, Lyon (France).