Journal of Safety Research 90 (2024) 350–370

Contents lists available at ScienceDirect

Journal of Safety Research

journal homepage: www.elsevier.com/locate/jsr

Meta-analysis of the safety effect of electronic stability control

A.E. af Wåhlberg [*], L. Dorn

Cranfield University, Cranfield, Bedfordshire MK43 0AL, UK

ARTICLE INFO

Keywords:
Electronic stability control
Induced exposure
Loss of control
Roll-over
Advanced driver assistance system

ABSTRACT

Objective: Electronic Stability Control (ESC) is a standard feature on most modern cars, due to its reported efficiency to reduce the number of crashes of several types. However, empirical studies of safety effects of ESC for passenger vehicles have not considered some methodological problems that might have inflated the effects. This includes self-selection of drivers who buy/use ESC and behavioral adaptation to the system over long time periods, but also the dominant method of induced exposure. This study aimed to investigate whether such methodological problems might have influenced the results. *Method:* A *meta*-analysis was undertaken to investigate whether there are systematic differences between published studies. Moderators tested included when the study was undertaken, the type of vehicle studied, the percent ESC in the sample, size of sample, the length of the study, whether matched or un-matched vehicles were studied, whether induced exposure was used, and two variants of types of crashes used as controls. *Results:* The effects found ranged from 38% to 75% reduction of crashes for the main targets of singles, running off road and rollover crashes. However, these effects were heterogeneous, and differed depending on the methods used. Most importantly, information that could have allowed more precise analyses of the moderators were missing in most publications. *Conclusions:* Although average effects were large and in agreement with previous *meta*-analyses, heterogeneity of the data was large, and lack of information about important moderators means that firm conclusions about what kind of mechanisms were influencing the effects cannot be drawn. The available data on ESC efficiency are not unanimous, and further investigations into the effects of ESC on safety using different methodologies are warranted.

1. Introduction

1.1. Increasing traffic safety by in-vehicle technology

Within traffic safety technology, few inventions are considered as successful as Electronic Stability Control (ESC). This kind of system stabilizes vehicles when they stray from their intended paths, or when their wheels are spinning, by applying the brakes separately for each wheel and adjusting engine torque. The result is a strong improvement in stability.

As ESC became increasingly common in the vehicle population, empirical evaluations of its effectiveness were published, and it was reported that vehicles with ESC had much fewer crashes than those without (40% to 50% for singles and some other categories; Erke, 2008). It was therefore hailed as an extremely effective intervention for vehicle safety (e.g., Fach & Ockel, 2009; Fitzharris, 2020; Fitzharris, Scully & Newstead, 2010; German Insurance Association, 2014; Krafft, Kullgren, Lie & Tingvall, 2009; NHTSA, 2011).

In this paper, it will be argued that there are reasons to believe that ESC is not quite as effective as reported and believed. It will be pointed out that the empirical investigations of the effects of ESC contain at least five problematic methodological features, which appear not to have been investigated. Furthermore, it will be argued that ESC research studies differ concerning some aspects of methodology, and that the resulting effect sizes therefore are heterogeneous and are difficult to interpret as estimates of the population effect.

1.2. Meta-analysis; aims and methodology

To support the claim of inflation in estimates of ESC safety benefit, a *meta*-analysis of the available literature on ESC effects on crashes was undertaken. Meta-analysis is the quantitative summary of available research on a specific topic. The main goal is usually to estimate the population effect size (in the case of ESC, the percent reduction in crashes) by averaging effects over studies. However, summarizing published research is almost never a straightforward affair, and some

---

[*] Corresponding author.
E-mail address: aw@empirica.pm (A.E. af Wåhlberg).

https://doi.org/10.1016/j.jsr.2024.07.004
Received 15 September 2023; Received in revised form 4 April 2024; Accepted 22 July 2024

important *meta*-analytic concepts will therefore be described here, to facilitate understanding of the analysis in this paper. Effect sizes (Cohen's d, Pearson r, odds ratios and several others) should only be averaged if they can be said to be drawn from the same population of effects, that is from studies that have investigated the same problem using the same methodology (Field & Gillett, 2010). Still, even under such circumstances, the effects will differ quite a lot due to random factors (Hunter & Schmidt, 1990), and smaller studies will tend to yield more disparate results. The degree of variation between studies can be quantified, and if it exceeds certain values, it can be suspected that the effects are not from the same population, and the data are said to be heterogeneous (Higgins & Thompson, 2002). Unfortunately, some tests of heterogeneity have low power (the ability to detect an actual effect), as they are significance-based, and *meta*-analytic samples are usually rather small. Outlying effects are also a common problem in *meta*-analysis (Viechtbauer & Cheung, 2010), and a special case of heterogeneity. This includes those that are extremes only in bi-variate associations (Ben-Gali, 2005), although this is seldom discussed. As with subject-level data, outliers can be suspected to be due to errors in the research process and might reduce otherwise strong associations to trivial values (or the opposite).

It is important to know that different methods often yield different effects when applied to the same problem, meaning that the effects are to some degree created by the method. Such a systematic difference is called a moderator, which can be controlled for statistically, or handled by calculating separate averages, for example for different age groups. When averaging effects, these are usually weighted by sample size to create equality for results from large and small studies.

Meta-analysis often suffers from a statistical power problem, because the number of effects (k) that can be included is small. This means that (even substantial) effects that are present in the population often cannot be detected. It is therefore very important to try to include as many samples as possible, which might come into conflict with the goal of only averaging samples from the same population of effects.

1.3. Induced exposure; assumptions and problems

The central problem of estimating effects of traffic safety interventions like ESC is the lack of exposure data for light vehicles. Without such data, number of crashes for vehicles with and without ESC cannot be compared, because any difference might be due to a difference in exposure to risk, and not the intervention as such.

The dominating solution for this problem is called induced exposure, which is a variant of case-control methodology. When actual exposure to risk of accident is not known, this can be estimated using another type of accident as the measure of exposure (Haight, 1973). To this method was added the assumption that drivers involved in non-culpable accidents are a random sample of the driving population (often called quasi-induced exposure; Stamatiadis & Deacon, 1997) and their characteristics could be assumed to be like those of the total population. Any difference between this group and the case group is therefore interpreted as a factor in accident causation. However, there are some inherent weaknesses of the induced exposure method, which do not seem to have been investigated in any detail.

The central problem of induced exposure is what kind of crash to use as denominator. The assumption of quasi-induced exposure is that drivers not responsible for their crashes are a random sample of the population (Lyles, Stamatiadis, & Lighthizer, 1991). This is reasonable, but it leads to a methodological problem; how do we know who is not culpable? The standard solution is to accept the judgment of police officers, but assigning blame is a highly subjective process, and differ depending on who makes the decision (af Wåhlberg, 2009a). In induced exposure studies, whether the judgment is correct or not is not tested in an objective way (Dorn and af Wåhlberg, 2018; af Wåhlberg & Dorn, 2007; 2019), and the degree of error will therefore not be known. In general, misclassification of culpability leads to under-estimation of effects (af Wåhlberg, 2018).

1.4. ESC study methodology

The usual method for tests of ESC effectiveness is to sample data from a crash database, according to which type of crash is believed to be influenced (target/case crashes), and which is not (non-target/control), which is the method of induced exposure. The vehicles involved in such crashes are then checked for ESC installation, yielding four different groups, the numbers of each which go into the calculation of an odds ratio (see Table 1). The formula used is (C1/C2)/(C3/C4) = Odds Ratio (OR), where 1 is no effect, and < 1 is a reduction in risk in C1 (in this case the use of ESC in certain situations). Odds ratios can then easily be converted into percent change (for example, OR 0.5 equals a 50% reduction, 1.5 a 50% increase).

However, induced exposure methods are not exactly the same between ESC studies, because although the assumptions are similar, how they have been operationalized vary strongly (Scully & Newstead, 2007). This can be seen in Tables 2-3, where non-target as well as target crashes vary between reports. The results of using different non-target crashes in safety calculations is not well known, as few studies have compared different non-targets (but see Jiang, Lyles, & Guo, 2014; Keall & Newstead, 2009; af Wåhlberg, 2009b).

ESC studies have typically used rear-end crashes as a measure of exposure, which can rather safely be assumed to be non-culpable for the struck party. However, some authors have included both the striking and the struck vehicle, while others have only used the struck vehicle. In many instances, the distinction is not made at all, and it remains unclear what crashes were included (as noted by Scully & Newstead, 2007).

Two slightly different methods of sampling vehicles have been used in ESC research. The first is to match vehicles, i.e., to restrict the sampling to makes and models that have been produced with and without active safety in roughly the same time period. This method strongly restricts the number of available vehicles, but it also controls for many possible differences in drivers, environments, and car characteristics. The other variant is to compare all available vehicles (ESC-equipped versus all others, which will be called non-matching). Matching of subjects is the traditional alternative to random sampling in social science, with the goal of minimizing error variance in the data. Less variance also means stronger effects in a sample, because effect sizes usually use some kind of ratio between the differences in means divided by the variation. Non-matched samples in ESC research can therefore be expected to have more variance in between themselves and smaller effects, because random factors will influence the effects to a higher degree as compared to matched samples.

Neither of these sampling methods controls for differences in buying decisions in relation to ESC, i.e., although the cars may be very similar, the drivers may be different. Also, changes in behavior might happen over long time periods that are not captured by these methods. These problems are further discussed in separate sections below.

Some researchers have split light vehicles into passenger cars and somewhat larger ones with a higher center of gravity, often called Light Trucks and Vans (LTV). The latter are believed to yield larger effects for ESC, as they are less stable to begin with.

All such methodological differences between studies need to be tested as moderators of effects.

Table 1
Odds ratio calculation principle with ESC as example.

	Case crash (e.g., rollover)	Control crash (e.g., rear-end)
Case vehicle (ESC)	C1	C2
Control vehicle (no ESC)	C3	C4

Table 2
Empirical studies of effects of Electronic Stability Control (Active Stability Control, Automotive Stability Management System, Dynamic Stability Control, Electronic Stability Program, Vehicle Stability Control etc) on crash involvement. Before-after comparison means the same model of car is compared when it has ESC and when not. This means that the ESC-equipped vehicles are usually a few years younger than those without ESC. The size of the study is given as the maximum number of crashes, while lower numbers will have been used for specific crash types. Most studies used induced exposure, calculated odds ratios, and transformed these to percent values. To save space, only values for a type of calculation which was common among the studies have been included in the Table.

Study	Population	Country, state	Data source	Method	Number of crashes in study (ESC/control)	Time period for crashes	Effects for specific crash types	Effects on all types of crashes (minus non-control) for this type of vehicle	Comments
Aga & Okada, 2003	Passenger cars	Japan	Police records	Before-after comparison of the same/similar models. Accidents per vehicle-years.	863/2058	1994–2001	−35% (single) −50% (single, severe damage) −35% (single, injury) −30% (head-on) −40% (head-on, severe damage) −35% (head-on, injury)	–	390 k vehicle years (ESC), 980 k (without). No significance or confidence limits reported. Three Toyota models. Only single and single injury values used in *meta*-analysis, as the rest were not comparable.
Bahouth, 2005	Passenger cars and LTV	US, Maryland	Police records	Before-after comparison of the same/similar models. Induced exposure. Control: Rear-end.	479/1320	1998–2002	−27.1% (multi-vehicle frontal) −74.9% (single)	–	Excluded results due to overlap with Dang, 2007: Florida, Illinois, Missouri.
		US, Texas			1288/4081		−11.5% (multi-vehicle frontal) −60.9% (single)	–	
		US, Utah			183/658		−9.8% (multi-vehicle frontal) −47.9% (single)	–	
Bahouth, 2006	Passenger cars and LTV	Florida	Police records	Before-after comparison of the same/similar models. Induced exposure. Control: Rear-end	3274/6374	1998–2003	–	−54% (fatal, serious injury) (−25% light injury)	This study to a large degree used the same state data as Dang, 2007, but the crash variables studied were different. The results therefore do not turn up twice in the same calculations. Results from FARS data excluded due to overlap with Dang, 2007. Number of crashes within parentheses excluded, as the the states did not match between tables.
		Illinois			4510/5351		–	−59% (fatal, serious injury) (−18% light injury)	
		US, Kansas			330/455		–	−42% (fatal, serious injury) (−51% light injury)	
		US, Kentucky			(1283/ 2263)		–	−67% (fatal, serious injury) (light injury)	
		US, Maryland			(966/ 1549)		–	−31% (fatal, serious injury)	

(*continued on next page*)

Table 2 (*continued*)

Study	Population	Country, state	Data source	Method	Number of crashes in study (ESC/control)	Time period for crashes	Effects for specific crash types	Effects on all types of crashes (minus non-control) for this type of vehicle	Comments
								(light injury)	
		US, Missouri			(120/249)		–	−58% (fatal, serious injury) (light injury)	
		US, North Carolina			(3077/4803)		–	−41% (fatal, serious injury) −(light injury)	
		US, Pennsylvania			(795/1313)		–	−63% (fatal, serious injury) (light injury)	
		US, Texas			(1402/4282)		–	−61% (fatal, serious injury) (light injury)	
		US, Utah			(240/710)		–	−58% (fatal, serious injury) (light injury)	
Chouinard & Lécuyer, 2011	Passenger cars and LTV	Canada	National Collision Database	Comparison between cars with and without ESC. Induced exposure. Control: Uncertain. Probably all non-ESC-sensitive crashes.	17 968/1 126 205	2000–2005	−41.1% (all ESC-sensitive; single, loss of control, ran-off-road, swerving, skidding) −18.6% (single) –23.2% (multi-vehicle) −54.8% (all ESC-sensitive injury; single, loss of control, ran-off-road, swerving, skidding) −49.3 (single, injury) −28.4% (multi-vehicle, injury)	−9.7% (all) −11.7% (injury)	The overall effects for ESC-sensitive crashes were not used (all and injury), as they did not correspond to any effects in other studies.
Dang, (2004); 2007	Passenger cars, only Mercedes	US, California	Police records	Before-after comparison of the same/similar models. Induced exposure. Control: Rear-end, low-	3941/3102	2001–2003	−66% (ran-off-road) −76% (rollover) –22% (culpable multi-vehicle)	–32%	FARS results not included, as they overlap with Kahane, 2014. The category ’Other single’ not used, as it apparently was a sub-group of all singles and therefore not

(continued on next page)

Table 2 (*continued*)

Study	Population	Country, state	Data source	Method	Number of crashes in study (ESC/control)	Time period for crashes	Effects for specific crash types	Effects on all types of crashes (minus non-control) for this type of vehicle	Comments
	Passenger cars	US, Florida		speed, non-culpable multi-vehicle	9242/9174	1997–2003	−28% (ran-off-road) −43% (rollover) −13% (culpable multi-vehicle)	−12%	comparable to other results.
		US, Illinois			5598/8123	1997–2002	−53% (ran-off-road) −80% (rollover) −% (culpable multi-vehicle)	−13%	
		US, Kentucky			907/1529	1997–2002	−47% (ran-off-road) −73% (rollover) −10% (culpable multi-vehicle)	−10%	
		US, Missouri			1927/2662	1997–2003	−44% (ran-off-road) −82% (rollover) −8% (culpable multi-vehicle)	−17%	
		US, Pennsylvania			1403/2520	1997–2001, 2003	−40% (ran-off-road) −66% (rollover) +10% (culpable multi-vehicle)	−2%	
		US, Wisconsin			1255/1919	1997–2003	−34% (ran-off-road) −58% (rollover) −16% (culpable multi-vehicle)	−11%	
Dang, (2004); 2007	LTV, only Mercedes	US, California	Police records		898/198	2001–2003	−81% (run-off-road) −78% (rollover) −24% (culpable multi-vehicle)	−35%	
Dang, (2004); 2007	LTV	US, Florida			2776/5326	1997–2003	−66% (run-off-road) −78% (rollover) −10% (culpable multi-vehicle)	−13%	
Dang, (2004); 2007		US, Illinois			1949/2916	1997–2002	−80% (run-off-road) −87% (rollover) −28% (culpable	–22%	

(continued on next page)

Table 2 (*continued*)

Study	Population	Country, state	Data source	Method	Number of crashes in study (ESC/ control)	Time period for crashes	Effects for specific crash types	Effects on all types of crashes (minus non-control) for this type of vehicle	Comments
							multi-vehicle)		
Dang, (2004); 2007		US, Kentucky			394/959	1997–2002	−77% (run-off-road) −92% (rollover) −17% (culpable multi-vehicle)	−17%	
Dang, (2004); 2007		US, Missouri			626/1369	1997–2003	−80% (run-off-road) −100% (rollover) +1% (culpable multi-vehicle)	−16%	
Dang, (2004); 2007		US, Pennsylvania			476/935	1997–2001, 2003	−63% (run-off-road) −86% (rollover) −1% (culpable multi-vehicle)	–22%	
Dang, (2004); 2007		US, Wisconsin			511/1027	1997–2003	−68% (ran-off-road) −88% (rollover) −27% (culpable multi-vehicle)	−35%	
Fildes et al., 2013	Passenger cars and LTV	Australia	Police records	Comparison between cars with and without ESC. Induced exposure Control: Rear-end	1247/24 324	2001–2005	−21% (injury, single) +0.5% (multi-vehicle, injury)	–	Values for singles averaged over the conditions of wet/dry and speeds (<75 km/ h >) in Table 4. * High percentage of ESC, deleted from analysis.
		Finland			343/3649	2000–2008	−14% (single, injury) +7% (multi-vehicle, injury)	–	
		Italy			14 614/ 5034*	2008	−44% (single, injury) −14.5% (multi-vehicle, injury)	–	
		New Zealand			194/2828	2001–2005	−31% (single, injury) +4% (multi-vehicle, injury)		
		Sweden			4880/12 859	2003–2010	−54% single, injury) −24.5% (multi-vehicle, injury)		
		UK			7172/23 942	2002–2005	−18% (single, injury)		

(*continued on next page*)

Table 2 (*continued*)

Study	Population	Country, state	Data source	Method	Number of crashes in study (ESC/control)	Time period for crashes	Effects for specific crash types	Effects on all types of crashes (minus non-control) for this type of vehicle	Comments
							−3% (multi-vehicle, injury)		
		Average for all countries			28450/72636	2001–2008	–	−13% (injury)	
Green & Woodrooffe, 2006a	Passenger cars	US	FARS	Before-after comparison of the same/similar models. Induced exposure. Control: Multi-vehicle.	346/1100	1995–2003	−30.5% (single, fatal) −34.8% (ran-off-road, fatal) −39.7% (rollover, fatal)	–	Eight models of cars. Differences in vehicle age did not affect calculations. Further results for wet roads available.
Green & Woodrooffe, (2006b); 2006a	LTV				202/271		−49.5% (fatal single) −56.4% (fatal ran-off-road) −72.9% (fatal rollover)	–	
Green & Woodrooffe, 2006a	Passenger cars	US	GES	Before-after comparison of the same/similar models. Induced exposure Control: Rear-end	1087/2835	1995–2003	−54.5% (loss of control)	–	
Green & Woodrooffe, (2006b); 2006a	LTV	US	GES		627/1634		−70.3% (loss of control)	–	
Kahane, 2014	Passenger cars	US	FARS	Before-after comparison of the same/similar models. Induced exposure. Control: Non-culpable multi-vehicle	4303/7118	1994–2011	−59.5% (fatal, first-event rollover) −31.3% (fatal, single vehicle, without rollover) −16.1% (fatal, culpable-multi-vehicle)	–	58 passenger car and 34 LTV models.
	LTV				6411/14808		−74% (fatal, first-event rollover) −45.5% (fatal, single vehicle, without rollover) −16.1% (fatal, culpable-multi-vehicle)	–	
Kallan & Jermakian, 2008	LTV	US	Police records, NASS-GES	Comparison between cars with and without ESC. Induced exposure. Control: Unknown*	278/2561	2001–2006	−67%, (single, rollover)	–	Adjusted OR 0.33. Effect put into category Rollover. This study did not explicitly state a specific category of crashes as control, but probably used all crashes that were not single rollovers for this end.

(*continued on next page*)

Table 2 (*continued*)

Study	Population	Country, state	Data source	Method	Number of crashes in study (ESC/ control)	Time period for crashes	Effects for specific crash types	Effects on all types of crashes (minus non-control) for this type of vehicle	Comments
Keall & Newstead, 2021	Passenger cars and LTV	Australia and New Zealand	Police records	Comparison between cars with and without ESC. Rate of crashes per registered vehicle.	144 603/ 90 581	2008–2017	−64.0% (rollover)	−8.9% (all)	Values calculated for both types of vehicles from Table 2. Values for ESC fitting per state not given.
Koisaari, Kari, Vahlberg, Sihvola & Tervo, 2019	Passenger cars	Finland	Finnish Motor Insurer's Centre Finnish Crash Data Institute	Comparison between cars with and without ESC. Crashes/km.	8827/21 437 (injury)	2009–2013	−42% (multi-vehicle, injury) −62% (single, injury) −77% (ran-off-road, injury) −67% (multi-vehicle, fatal) −70% (single, fatal) −73% (ran-off-road, fatal)	−49% (all injury) −68% (all fatal)	Only injury-related and fatal at fault crashes, excluding pedestrian and cyclist involvement. Unadjusted values for rates per mileage used.
Kreiss, Schüler & Langwieder, 2005	Passenger cars	Germany	German Federal Statistical Office	Comparison between cars with and without ESC. Induced exposure. Control: Uncertain, probably several categories.	12 610/28 422	1998–2002	–32.4% (loss of control) −55.5% (loss of control?, fatal)	–	Some uncertainty about whether the values were for the specific loss of control category or all crashes. Values substantially higher after correction for misclassification.
Lie, 2012	Uncertain. Probably passenger cars.	Sweden	Swedish Transport Administration	Before-after comparison of the same/ similar models. Induced exposure. Control: Uncertain. Probably all non-ESC-sensitive crashes.	68/60	2004–2010	−74% (loss of control, fatal)	–	–
Lie, Tingvall, Krafft & Kullgren, (2004; 2005); 2006	Passenger cars	Sweden	Police records	Before-after comparison of the same/ similar models. Induced exposure. Control: Rear-end	1 942/8 242	1998–2004	−44.4% (single, serious/ fatal)	−16.7% (injury) −21.6% (serious/ fatal)	–
Lyckegaard, Hels & Bernhoft, 2015; (Hels, Lyckegaard & Berntoft, 2013)	Passenger cars?	Denmark	Police records	Comparison of vehicles with and without ESC. Induced exposure Control: All	3121/10 515	2004–2011	−60% (injury, single) −60% (fatal, single) −58% (severe	–	This study assumed that only single crashes are influenced by ESC, in contrast to most other studies. Number of crashes in the different categories estimated

(continued on next page)

Table 2 (*continued*)

Study	Population	Country, state	Data source	Method	Number of crashes in study (ESC/control)	Time period for crashes	Effects for specific crash types	Effects on all types of crashes (minus non-control) for this type of vehicle	Comments
				but single crashes			injury, single)		from Appendix AI. Single severe injury not used, as it was not comparable.
MacLennan et al., 2008	Passenger cars and LTV	US	Police records, GES	Comparison of vehicles with and without ESC. Induced exposure. Control: Multi-vehicle	8198/296 437	1996–2006	−38% (rollover) (risk ratio 0.62)	–	Values given in paper, although a risk ratio does not equal an odds ratio, and the percent estimate is therefore likely somewhat too small. However, see note about Sivinski (2011) below.
Page & Cuny, (2005); 2006	Car (one model)	France	French National injury accident census	Before-after comparison of the same/similar models. Induced exposure. Control: Rear-end, pedestrian, junctions, etc	93/495	2000–2003	−43% (injury, loss of control)	–	One model of car.
Riexinger, Sherony & Gabler, 2019	Passenger cars and LTV	US	NASS/CDS	Comparison of vehicles with and without ESC. Induced exposure. Control: Rear-end.	4944 total, estimated 584/4360	2006–2015	−51.4% (loss of control) −13.3%* (rollover)	–	* Outlying value, excluded.
Scully & Newstead, 2010	Passenger cars and LTV	Australia and New Zealand	Police records	Before-after comparison of the same/similar models. Induced exposure. Control: Rear-end.	24 235/ 308 298	2001–2008	−27.58% (single, all severities) –32.31% (single, driver injury) –22.32% (single, driver serious injury) −45.95% (rollover, all severities) −42.68% (rollover, driver injury) −25.43% (rollover, driver serious injury) +2.78% (multiple vehicles, all severities) −6.98% (multiple vehicles, driver injury) −5.01% (multiple vehicles, driver serious	−2.9% (all severities) −13.96% (driver injury) −11.82% (driver serious injury)	Unadjusted values used to be comparable to other results. Further results available for different types of vehicles etc. Five Australian states. CL not given for unadjusted values. Head on, single serious, rollover and multiple serious injury values not used, as they were not comparable.

(continued on next page)

Table 2 (*continued*)

Study	Population	Country, state	Data source	Method	Number of crashes in study (ESC/control)	Time period for crashes	Effects for specific crash types	Effects on all types of crashes (minus non-control) for this type of vehicle	Comments
							injury) −8.80% (head-on, all severities) −1.29% (head-on, driver injury) −8.78% (head-on, driver serious injury)		
Sivinski, 2011	Passenger cars	US	Police records, GES	Before-after comparison of the same/similar models. Induced exposure. Control: Rear-end, non-culpable multi-vehicle, low speed	8 040 for both, 40.7% ESC*	1997–2009	−34.8% (single) −49.7% (single, fatal) −3.4% (multi-vehicle) −8.4% (multi-vehicle, fatal) −65.3% (rollover) −56.4% (rollover, fatal)	−17.5% −31.8% (fatal)	The statistical method used is called risk ratio by the author but is actually an odds ratio. Weighting process not described. Results for FARS data not included here, as they overlap with Kahane, 2014. Values for several other crash types available. * Value for population from weights on p. 12. Unweighted values for effects used but weighted for CI.
	LTV						−54.5% (single) −54.6% (single, fatal) −2.1% (multi-vehicle) −7.7% (multi-vehicle, fatal) −69.1% (rollover) −74.2% (rollover, fatal)	−20.0% −30.4% (fatal)	
Strandroth, Rizzi, Olai, Lie & Tingvall, 2012	Passenger cars	Sweden	Police records	Type of comparison uncertain, assumed to be different vehicles. Induced exposure. Control: Rear-end	4790/7356	2003–2010	−52% (loss of control, injury) –23% (loss of control, fatal)	–	–
Thomas & Frampton, 2007; (Frampton & Thomas, 2007; Thomas, 2006a; 2006b; Weekes, Avery, Frampton & Thomas, 2009)	Passenger cars and LTV	UK	STATS19	Before-after comparison of the same/similar models. Induced exposure. Control: Crashes with various low-speed behaviours	10 475/41 656	2002–2005	−7% (injury) −11% (serious injury) −25% (fatal) –33% (skidding) −59% (rollover) −27% (single)	–	Non-target behaviours included reversing, parked, waiting, stopping, starting. Includes the data from Thomas, 2006a; 2006b. Injury crashes, without vulnerable road users,

Table 3
Studies and results on ESC not included in the present *meta*-analysis.

Study	Population	Country	Data source	Method	Number of crashes in study (ESC/ control)	Time period for crashes			Comments
Becker, Delaney & Newstead, 2007	?	Germany	?	?	?	?	?	−45% (Scully & Newstead, 2007)	Referenced in Scully & Newstead, 2007. Study not found.
Dang, (2004); 2007	LTV	US	FARS	Before-after comparison of same models. Induced exposure. Control: Rear-end, low-speed, non-culpable multi-vehicle	261/583	1997–2004	−63% (single, fatal) −70% (run-off-road, fatal) −88% (rollover, fatal)	−28% (fatal)	FARS data excluded due to overlap with Kahane, 2014.
	Passenger cars	US			642/676		−36% (single, fatal) −36% (run-off-road, fatal) −70% (rollover, fatal)	−14% (fatal)	
Farmer, (2004) 2006	Passenger cars and LTV	US	Police records	Comparison between cars with and without ESC.	11 969/19 851	2001–2003	−41% (single vehicle) −56% (single vehicle fatal) −79% (single vehicle rollover, fatal)	−2% (all) −5% (injury) −43% (fatal)	25 models of vehicles. Self-selection problems possible. Seems to assume equal exposure between equipped and non-equipped. Many more effects for specific categories available. Excluded due to large overlaps with Dang, 2007.
Farmer, 2010	Passenger cars and LTV	US	FARS	Crash rates per registered vehicle.	–	1999–2008	−49% (fatal single)	–33% (fatal) −20% (fatal, multi-vehicle)	Values from Blower, 2014. Excluded due to overlap with Kahane, 2014. Study not found.
Fennel, 2003	?	Germany	Time series	?	?	?	−12% (rollover)	−15%	Values from Burton, Delaney, Newstead, Logan & Fildes, 2004. The full reference was not given. Study not found.
Grömping, Weimann & Menzler, 2004	?	Germany	GIDAS	Comparison between cars with and without ESC. Induced exposure Control: Uncertain. Probably all non-ESC-sensitive crashes.	6211?	1994–2003	–	–	Only adjusted values for specific scenarios available.
Johnson & Gabler, 2015	Passenger cars	US	FARS	Before-after comparison of the same/similar models.	1 016/1 514	1997–2011	−47.3% (fatal, barrier)	–	Excluded due to overlap with Kahane, 2014.
Johnson & Gabler, 2015	LTV	US	FARS	Before-after comparison of the same/similar models.	947/2 056	1997–2011	−39.7% (fatal, barrier)	–	
Padmanaban, 2007	Passenger cars LTV	US	FARS	Before-after comparison of the same/similar models. Induced exposure. Control:	1966 vehicles 1833 vehicles	1995–2004?	−52.5% (rollovers) −55.6 (rollovers)	– –	Different time periods given in different parts of the text. Excluded due to overlap with Dang, 2007 and Kahane, 2014.

(*continued on next page*)

Table 3 (*continued*)

Study	Population	Country	Data source	Method	Number of crashes in study (ESC/control)	Time period for crashes			Comments
				stationary, slow, rear-end, non-culpable multi-vehicle					
Scully & Newstead, 2007; 2008	Cars and LTV	Australia and New Zealand	Police records	Comparison of vehicles with and without ESC. Induced exposure. Control: Rear-end, both vehicles	7 699/203 186	2001–2005	−27.65% (single, police-reported) −30.39% (single, injury) +11.94 (multi-vehicle) −4.05 (multi-vehicle, injury)	+5.18% (police-reported) −10.73 (injury)	Unadjusted values. Excluded due to data overlap with Scully & Newstead, 2010. Five Australian states.
Tingvall, Krafft, Kullgren & Lie, 2003	–	Sweden	–	– Induced exposure. Control: Rear-end.	442/1967	2000–2002	–	–22.1% (wet roads) −38.2% (icy roads)	Values from Burton, Delaney, Newstead, Logan & Fildes, 2004. Reference not found. Probably an early version of later papers with the same name by the same group.
Unselt, Breuer, Eckstein & Frank,	Mercedes cars	Germany	Federal Statistical office	Trends in loss of control crashes for different makes of cars.	>2 000 000	1998–2002	−40% (loss of control)	–	The percent reduction refers to a decrease in the fraction of crashes that were of the loss of control type for Mercedes, while the trend for other makers was not considered, although it was said to be slightly positive too. Calculation not comparable to other data.
Yannis, Papadimitriou, Dupont & Martensen, 2010	?	Several European countries	Police records etc, FAI database	Comparison between cars with and without ESC.	?	?	−65% (fatal, occupants in cars)	–	The group sampled was not comparable to those of other studies.

1.5. *Esc-sensitive crashes*

Many studies on ESC have defined their target crashes very narrowly according to where the effect is believed to take place ('relevant for ESC'). This means that very high values for crash reduction can be achieved, because the population targeted is very limited. Percentages that might look impressive can reflect small reductions in the total number of crashes (and very few researchers have calculated the effect in percent of all crashes, e.g., Sivinski, 2011). This problem can be illustrated by the study reported by Chouinard and Lécuyer (2011), where values for both 'ESC-sensitive' and all crashes in a category were given. For example, the reduction for 'All crashes' was 9.7%, but 41.1% for the 23.7% of crashes that the authors had considered to be ESC-sensitive. In the absence of a clearly reported method such as that of Chouinard and Lécuyer, results might be misleading, given that 'x percent of singles' may refer to 'x percent of a special sub-sample of singles.'.

For the *meta*-analyst, this feature becomes problematic, as different researchers may have targeted different populations of crashes, by creating different outcome variables, which adds heterogeneity to the data. This means that the results are difficult to average, and to interpret.

If crashes are very narrowly defined as 'ESC-relevant,' the effects found are likely to be higher than with a broader definition. Although it was not possible in the present study to code for definitions used, increasing narrowness of the definitions used will lead to decreasing sample sizes. Unfortunately, many other factors would also influence sample size, making it difficult to discern this effect with an analysis of the association between effects and sample sizes. However, in the absence of a better alternative, this analysis was undertaken in the present study. It can be noted that this kind of analysis is often used to detect publication bias (Ioannidis, Munafò, Fusar-Poli, Nosek, & David, 2014), which can also explain such an association, if found.

1.6. *Behavioral adaptation*

One shortcoming of all ESC studies is that they do not consider how long a driver has been driving a vehicle with this technology, and therefore whether behavior might change during or after the study. In traffic safety research, it is well known that interventions can lead to counter-productive behavior, which lessens their impact (Branas & Knudson, 2001; Erke, 2009; Herms, 1972; Hollingworth, Harper, & Hamer, 2015; Peltzman, 1975; Phillips, Fyhri, & Sagberg, 2012; Rudin-Brown & Parker, 2004; Rumar, Berggrund, Jernberg, & Ytterbom, 1976;

Table 4
Descriptive values for sample sizes and methodological factors. k = 57. The sample numbers refer to the total sample extracted, not the sub-categories of crashes.

	Mean	Max	Min	Std
Study period midpoint	2002.15	2013	1999.5	3.1
Number of ESC in sample	6 320	144 603	68	19 530
Number of non-ESC in sample	38 932	1 144 173	60	159 488
Percent ESC in sample	31.6%	81.9%	1.5%	15.8

Thulin, 2007), including reactions to automated vehicles (Soni, Reddy, Tsapi, van Arem, & Farah, 2022). Although little is known about how long it takes for drivers to change their driving behavior in response to new safety technology (see theories by Hedlund, 2000; Rudin-Brown & Noy, 2002; Summala, 1997; Wilde, 1982), or why this seems to happen only for some interventions, it is well established that it does happen at times (Rudin-Brown & Jamson, 2013). This includes the precursor of ESC, anti-lock brakes (Aschenbrenner & Biehl, 1994; Sagberg, Fosser, & Sätermo, 1997), which is currently considered by some to have no safety effect (e.g., Bayly, Fildes, Regan, & Young, 2007; Burton, Delaney, Newstead, Logan, & Fildes, 2004; Kahane & Dang, 2009; Krafft, Kullgren, Lie, & Tingvall, 2009), although no *meta*-analysis seems to have been published on the topic (but see Broughton & Baughan, 2002; Farmer, 2001; Hertz et al., 1998, for some results). It would seem to be peculiar that this very similar system (in terms of being a system that adjusts the control input from the driver) should yield so very different results from ESC, but the reasons for this have not been investigated (Vaa, Penttinen & Spyropoulou, 2007).

For induced exposure, C1 is the target (see Table 1), where the effect is believed to take place. However, a decrease in odds ratio could be also due to an increase in C2. This could result from behavioral adaptation, where drivers become more careless and cause other types of crashes (Rudin-Brown, Jenkins, Whitehead, & Burns, 2009; Vadeby, Wiklund, & Forward, 2011), or simply become more efficient in their braking, leading to more rear-end crashes. In principle, a low odds ratio could also be due to changes in C3 and/or C4, although this is extremely improbable. However, Bahouth (2005) discussed apparent changes in reported proportions of rear-end crashes with age of vehicles, and similar effects, and Erke (2009) found increases in rear-ending at traffic lights with surveillance cameras.

The weak point of this kind of method is therefore that behavioral adaptation might increase C2, or other types of crashes that have not been considered at all, and possibly not until the study period has ended. The total safety effect of ESC may therefore be smaller than calculated using the induced exposure method (and Erke, 2008, pointed out that some effects were suspiciously large as compared to what was expected).

In a *meta*-analysis of ESC, an effect of behavioral adaptation could show up as decreased effects in studies over longer time periods and may coincide with larger samples, as more vehicles with ESC become available for analysis.

1.7. Self-selection of drivers

Self-selection of drivers in crash data has rarely been discussed in the road safety literature (for an exception, see Thomas & Frampton, 2007). It will here be proposed that safety-minded drivers may tend to buy cars with extra safety features (e.g., Girasek & Taylor, 2010; Koppel, Clark, Hoareau, Charlton, & Newstead 2013; Koppel, Charlton, Fildes, & Fitzharris, 2008), and that this difference between drivers adds to the true effect of ESC in any sample. Furthermore, this effect should be most noticeable when ESC is uncommon. In research and traffic safety statistics, this means that the reduction in crashes will be diluted over time as market penetration increases. In *meta*-analysis of ESC studies, this dilution effect could be noticeable as a trend towards smaller effects in samples over time. However, this trend could be countered by the selection of only a small number of vehicle models to study and other changes in methods.

It is usually, but implicitly, assumed by researchers that the effect of ESC will continue over time, in the sense that when the cars with ESC are sold on to other owners, they will provide the same protection. However, it could also be the case that drivers who buy older cars have different behaviors and do not equally benefit by the inclusion of ESC. If so, the effect of ESC would wane in the population as time goes by and market penetration increases.

A contrary hypothesis was suggested by a reviewer of this paper; at the time of the early evaluation studies, ESC was only available on a few luxury brands, and the self-selection problem would then have been smaller, because these buyers would mainly be interested in the brand itself, not the safety features. This would lead to increasing effects over time.

Table 5
Descriptive statistics for the effects data in Table 2, unweighted. Negative values indicate a reduction in crashes for ESC-equipped vehicles.

Crash type	k	Mean percent reduction	Max percent reduction	Min percent reduction	Std
All fatal	5	−35.4	−68.0	−21.6	18.7
Fatal singles*	9	−48.4	−70.0	−30.5	12.6
Fatal ran-off-road	3	−54.7	−73.0	−34.8	19.2
Fatal rollovers	6	−62.8	−74.2	−39.7	13.7
Fatal multi-vehicle	5	−23.1	−67.0	−7.7	24.9
Fatal loss of control	3	−50.8	−74.0	−23.0	25.8
Severe injury*	12	−46.4	−67.0	−11.0	19.4
All injury*	16	−30.9	−51.0	−7.0	13.7
Injury singles	11	−38.2	−62.0	−14.0	16.8
Injury ran-off-road	1	−77.0	−77.0	−77.0	–
Injury rollovers	1	−42.7	−42.7	−42.7	–
Injury multi-vehicle	9	−12.0	−42.0	7.0	16.6
Injury loss of control	2	−47.5	−52.0	−43.0	6.4
Singles	8	−43.3	−74.9	−18.6	19.5
Loss of control	5	−48.3	−70.3	−32.4	16.0
Ran-off-road*	14	−59.1	−81.0	−28.0	18.1
Rollovers*	22	−68.7	−100.0	−13.3	20.1
Multi-vehicle*	21	−12.1	−28.0	10.0	10.8
All police-reported*	19	−17.0	−35.0	−2.0	9.4

* Variables included in test for similarity between random effects model and sample-weighted and raw averages.

Table 6
Descriptive values for continuous moderators. k = 57. The sample numbers refer to the total sample extracted, not the sub-categories of crashes.

	Mean	Max	Min	Std
Study period mean time point (calendar year)	2002.15	2013	1999.5	3.1
Percent ESC in sample	31.6%	81.9%	1.5%	15.8
Size of sample (crashes)	44 938	1 144 173	128	162 172
Length of study period (years)	7.1	18	1	2.99

Table 7
Results from a random effects *meta*-analysis for studies where confidence intervals or raw data were available. These results are dominated by the studies of Bahouth (2006) and Dang (2007). Shown are the number of samples for each crash category, the weighted average odds ratio with confidence intervals, two measures of heterogeneity of the data (Q and I^2), a regression test of publication bias and results for three moderator variables from a *meta*-regression. − = no variance in the moderator.

Crash type	k	Odds ratio (CI)	Q	I^2	Publication bias (Egger's)	Size of sample	Percent ESC in sample	Type of vehicles
Effect on fatal singles	8	0.555 (0.485–0.636)	20.6**	66.1%	ns	ns	ns	*
Effect on severe injury crashes	10	0.474 (0.394–0.570)	5.6	0	ns	–	–	–
Effect on all injury	11	0.683 (0.617–0.756)	23.4**	57.3%	*	*	–	–
Effect on ran-off-road	14	0.396 (0.319–0.490)	99.4***	86.9%	**	ns	ns	*
Effect on rollovers	20	0.303 (0.228–0.403)	121.9*	84.4%	ns	ns	ns	*
Effect on all multi-vehicle	20	0.872 (0.826–0.921)	69.8***	72.7%	ns	*	*	ns
Effect on all police-reported crashes	16	0.820 (0.779–0.864)	38.4***	61.0%	ns	*	*	*

*p < 0.05, **p < 0.01, ***p < 0.001.

1.8. Hypotheses and analyses

The aims of the present study were to estimate average effects of ESC on different types of crashes for all available studies, but also to investigate possible moderators in these studies. These can be summarized as follows.

1) Self-selection of drivers might inflate the effects when comparing cars with and without ESC, but this effect would decrease as ESC became more common, leading to smaller effects over time. However, if buyers are more motivated by adherence to a certain brand of car, effects will increase over time.

2) Different types of vehicles have different effects. In the available data, vehicles with higher center of gravity (coded as LTV) were expected to yield larger effects.

3) The use of different types of crashes as the denominator (control) in calculations were expected to have an effect, but based on the data available, the direction of the effect could not be predicted.

4) The percentage of vehicles with ESC in the samples was predicted to be associated with the effects reported, but as the methodology has differed quite a lot between studies concerning how vehicles have been sampled, it was not possible to predict the direction of the effect.

5) Sample sizes were expected to correlate with effect sizes. As market penetration of ESC increased, sample sizes will have increased, and effects decreased.

6) The use of matched or non-matched vehicles will yield larger effects for the former, as matched vehicles should have less error variance.

However, several of these predictions were expected to interact, and the final effects could not be predicted. Although at least three *meta*-analyses of crash effect studies have been published (Erke, 2008; Høye, 2011; Oh, Youn, Jeong, & Oh, 2017), only the moderating effects of the type of vehicle and matching have been tested before. Also, Høye (2011) tested for differences due to controlling statistically for differences in driver characteristics, which is here replaced by other tests.

2. Method

2.1. General methodology

The current study was limited to light four-wheel vehicles as studies on heavy truck ESC usually use a different method for handling exposure and are less prone to self-selection issues. Also, trucks have very different handling characteristics, and it can be assumed that effects will differ versus those for light vehicles. The aims were to estimate the population effects on some different crash types and investigate possible method effects. A *meta*-analytic approach was used, but with some important differences to standard analyses of this kind, due to specific data characteristics. Most importantly, the number of samples (k) was increased by averaging values within samples.

Table 8
Meta-regression random effects model with the average of standardized effects per sample as dependent variable and all moderators as predictors. Only significant moderators retained in the model. k = 57.

Predictors	Coefficient	Standard error	95% lower	95% upper	Z-value
Intercept	12.2994	5.6104	1.3032	23.2956	2.19*
Study period mean time point	−0.0062	0.0028	−0.0117	−0.0007	−2.21*
Matched/non-matched vehicles	−0.0743	0.0191	−0.1117	−0.0369	−3.89***
Type of vehicle	0.0218	0.0090	0.0042	0.0394	2.43**

*p < 0.05, **p < 0.01, ***p < 0.001.

Table 9

Pearson correlations between the average and standardised effect sizes per sample and (more or less) continuous moderators. k = 57. Negative correlations between the effect variables and moderators mean an increase with values of the moderators.

Variable	Average of standardised effects	Study period mean time point	Type of vehicle	Percent ESC in sample	Size of sample	Length of study period
Average of effects	0.769***	0.095	0.236	−0.305*	0.210	0.020
Average of standardised effects		0.101	0.304*	−0.372**	0.137	0.230
Study period mean time point			0.080	0.047	0.119	0.119
Type of vehicle				0.061	−0.023	−0.036
Percent ESC in sample					−0.312*	−0.142
Size of sample						0.032

*p < 0.05, **p < 0.01.

2.2. Reference search

The Rosap collection of the U.S. National Transportation Library, Scopus and the Article database of the Australasian College of Road Safety were searched with the term "electronic stability control." Google Scholar was searched for the same term, but as this engine typically yields thousands of hits, only the first 300 titles were scanned.

Several reviews and *meta*-analyses on ESC have been published, and were used as sources of references (Bayly, Fildes, Regan, & Young, 2007; Blower, 2014; Burton, Delaney, Newstead, Logan, & Fildes, 2004; Erke, 2008; Ferguson, 2007; Høye, 2011; Wang, Zhong, Ma, Abdel-Aty, & Park, 2020). These reviews included references to some publications or findings by companies or organizations that cannot be sourced today (e.g., Bayley, Fildes, Regan, & Young, 2007) and suggests that more data are available than could be gathered for the present study.

Finally, the reference lists of all ESC papers found were searched, as were the citations in Google Scholar for the *meta*-analyses by Erke (2008) and Høye (2011) and the review by Ferguson (2007). A PRISMA flowchart describing this process can be found in the Appendix.

2.3. Inclusion of studies and effects

Empirical studies of the effects of ESC on crash involvement were included if they had values for risk reduction based in differences in crash numbers for vehicles with and without ESC. All reports found were in English, but some Swedish, German, Danish and Finnish studies were reviewed before being excluded as they contained no effect sizes or were duplicates of studies published in English. Duplicate works between and within researchers were identified and the older ones excluded. This feature was especially problematic for the U.S. data, which has been analyzed by many researchers, and where overlap in data and calculations between publications is often obscure.

Most studies reported effects for several crash types, but these were not the same between reports, as beliefs about what kind of crashes could be influenced by ESC differed between researchers. To create effect variables with reasonable numbers of values, it was decided to include only commonly used ones. Therefore, specific effects that were seldom reported (e.g., skidding crashes) were excluded.

Table 10

Differences in standardised averaged effect between dichotomous groups of different methodologies applied. These were using the same kind of vehicle with and without ESC versus all kinds, induced exposure versus other exposure controls, and two variants of induced exposure control crashes.

Variable	t	k
Same vehicles	−2.43**	41/16
Method: Induced exposure	−0.79	54/3
Controls: Several crash types	−0.38	28/29
Controls: Rear-end	0.03	26/31

2.4. Coding and moderators

Many studies reported values for sub-categories of general categories of crashes. When such instances were encountered, values from the wider range of conditions were used (i.e., 'rollover' was chosen before 'rollover due to loss of control'). Categories of effects were selected to be as similar as possible between studies. If a category had been split into several sub-categories and no overall figure was given in a study, the effects were averaged (e.g., wet and dry surfaces; Fildes et al., 2013, see further Table 2).

To estimate percent ESC in the population of vehicles from which the sample of crashes was drawn, the number of vehicles with and without ESC in the total sample was used. Some problems with this method can be noted; if the number of models included was restricted to those with early adoption of ESC, the percentage would be much larger than that of the population. As an alternative, the numbers of the control group could have been used. However, these latter values were seldom available. Therefore, a test using values for the overall samples and the controls in Dang (2007) was run. The percent ESC calculated from the total sample and from the control sample correlated 0.99. Although the means differed by a few percent, the relative standing of a sample on percent ESC was captured well by the total sample.

As passenger cars and LTVs could be expected to yield different effect sizes (vehicles with a higher center of gravity are expected to have more benefits from ESC), the type of sample was included as a moderator with three levels; LTV (including SUVs), mixed, and passenger cars.

Similarly, there were two different methods used for extracting samples; trying to control for differences between models/makes by matching the same model before and after introduction of ESC and collecting larger numbers by using all vehicles. However, some researchers appear to have accepted similar models as the same, for example Audi A2 could be matched with Audi A3 (Lie, Tingvall, Krafft, & Kullgren, 2006), Still, a difference can be expected between matching and non-matching, although it is not possible to say the direction of the effect. A matching procedure should in principle yield larger effects, but a self-selection effect would be larger for different vehicles.

The mid-time year of each sample was calculated and entered as a moderator, as it was believed that effects could wane over time.

Finally, the type of control crash used was dichotomously coded for rear-end and several crash types in two variables, to account for those that did not fit into these categories (i.e., Aga & Okada, 2003; Koisaari, Kari, Vahlberg, Sihvola, & Tervo, 2019). Thus, there was one moderator variable for rear-end/not rear-end, and one variable for the control was a single type of crash, or several types. This resulted in a strong overlap between these variables, as rear-ends tended to be used as a single type of control crash.

2.5. Studies and data excluded

In the literature on effects of ESC, there are many papers cited as evidence. However, a scrutiny of these yielded several problems, and some papers were therefore not included here.

Table 11
The associations between effect sizes and percent case crashes of the total sample for the crash categories where k was more than ten.

Variable	k	Correlation
Ran-off-road	14	0.15
Rollovers	16	0.69**
Multi-vehicle	14	0.24
All police-reported	15	0.35

Table 3 shows studies and effects that were fully excluded because they could not be sourced, or that had partial overlap with other studies. If studies overlapped with previous studies by the same authors, the earlier publications are given within brackets in Table 2.

2.6. *Crash categories*

With regards to data coding, the meaning of an effect on all crashes was not always clear. The general meaning of reporting that there was an x percent reduction in crashes would be that there was this reduction in the total of every type of crash. However, some authors would at times use the phrase 'all accidents' to denote the target crash types, excluding non-target and non-used crashes. The latter would of course mean a much higher reduction value. In this analysis, 'all crashes' were all types of crashes but with control crashes excluded.

2.7. *Weighting of effects, handling of data and analysis*

In *meta*-analysis, weights are usually attached to each effect in the analysis, to give each individual response the same weight. In the present data, the weighting was a problematic issue, as few studies had provided the number of vehicles/crashes or confidence intervals for each effect size reported. To make it possible to utilize the full *meta*-data, two different methods could be used; weighting by the total number of crashes in each sample (i.e., not the sub-sample used for each effect calculation) or using no weights at all. Neither of these methods could be said to be satisfactory. A test was therefore undertaken to discern which one could be assumed to yield the most accurate results, as compared to data where confidence intervals or sample sizes were available.

For each of the effects where confidence intervals had been provided or could be calculated from raw data for each crash category, random effects *meta*-analyses were undertaken for each category with more than eight values (see Table 7). The results from these analyses were compared to weighting by total sample size, using the method recommended by Hunter and Schmidt (1990), and to a raw average, for the same effects. It was found that a raw average was very similar to the corresponding random effects result in each crash group, differing at most 2.2%. Results for the sample-weighted analyses differed far more from the random effects ones. Also, Erke (2008) found that ESC effects did not differ with and without weights. Therefore, it was concluded that it would be acceptable to average effects without any weights.

The analysis was undertaken in three stages. First, a descriptive analysis without weights was undertaken to identify outliers and variables with enough data points for further analysis. Second, each crash type was analyzed separately for heterogeneity, dissemination bias, and moderator influences, using standard *meta*-analytic methods. This limited the data to samples where raw data or confidence intervals had been reported, and given the many different variables used in studies, the number of studies in each sample was very small.

Third, two different methods were used to pool all effects into a single one per sample to counter the power problem of the ESC data. First, all effects were averaged per sample and converted into correlations, which can be *meta*-analyzed with number of subjects as estimate of precision. Second, effects within each variable were standardized, whereafter they were averaged within each sample. This variable could not be *meta*-analyzed with weights, as it was expressed in units that are not effect sizes, but had the advantage of reducing the problem of the studies using different variables (crash categories) with expected different population effect sizes. This last analysis was best suited for moderator analysis, while the first and second were aimed at determining the average effects in the variables.

Meta-analytic analysis was undertaken in the Comprehensive Meta-Analysis (CMA) software, while descriptive statistics and correlations were calculated in Statistica.

The analyses included two tests of heterogeneity of effect sizes, Q and I^2 (Huedo-Medina, Sanchez-Meca, Marin-Martinez, & Botella, 2006). Such tests indicate whether effects vary more between samples than could be expected by random sampling effects. If excess variance is detected (heterogeneity), it can be suspected that these effects have been influenced by moderators. The two tests applied here work in different ways. While Q is based in significance testing and yields a result that is dependent both upon the amount of heterogeneity and the number of studies (k), I^2 estimates the amount of excess variance from 0 to 100%. It can be noted that both Erke (2008) and Høye (2011) used the Q test, and therefore had low probability of detecting heterogeneity in their ESC data.

A test for publication bias (Egger's) was also included. This works by testing whether small studies tend to have more positive findings than larger ones.

3. Results

3.1. *Errors, outliers, and uncertainties in data*

During the coding of the data, many instances of problematic data were encountered, which will be listed here. In Bahouth (2006), which included data from 10 U.S. states, it was found that the states listed

Table 12
The estimated changes in effect sizes with changes in moderators in the averaged dataset. All changes in effect size expressed as reductions. k = 57.

Moderator	Change in moderator	Change in effect size in percent units	Comment
Study period mean time point	1 year	0.45	Later studies had smaller effects
Type of vehicle	LTV/passengers cars	12.11	Passenger cars had smaller effects than LTV, but mixed cars/LTV even smaller
Percent ESC in sample	1% unit	0.29	Higher percent of ESC decreases effect
Size of sample	10 000 crashes	0.2	Larger samples had smaller effects
Length of study period	1 year	0.1	Longer study periods had smaller effects
Matched/non-matched vehicles	Type of comparison	14.05	Non-matched vehicles yielded smaller effects
Method: Induced exposure/other	Inducing or estimating exposure as miles or number of vehicles	5.34	Induced exposure had smaller effects than other methods (only three studies used other methods)
Controls: Several crash types/single type	Using only one or several types of crashes as controls	6.92	Using a single type yielded smaller effects
Controls: Rear-end crashes/other crashes	Using rear-ends or other crashes	8.04	Rear-end crashes yielded smaller effects

differed between Table 3(crashes) and Tables 4 and 5(effects). The number of crashes and the effects can therefore be associated with a certain state, or each other, but not with any certainty.

In Fildes et al. (2013), the percentage of cars fitted with ESC in the Italian sample (Table 2, p. 276) was suspiciously high (74%), as this should have been an all-vehicles sample, not matched models. No explanation for this could be found, as it would seem improbable that the Italian population of passenger cars at that time could have reached such a high market penetration for ESC.

In the study by Riexinger et al. (2019), two different values of effectiveness were given for rollover crashes, with the result for All rollovers substantially lower (13.3%) than the other (50.6%). The second category (Control loss rollovers) was probably a sub-category of the first, so for this to happen, the second category must have been carrying virtually all the effect. This means there was no effect on non-control loss rollovers, which is peculiar. Furthermore, the lower value was also a two standard deviations outlier on the rollover variable in the present study, where the unweighted average was 69% (see Table 5).

In Lie, Tingvall, Krafft, and Kullgren (2006), the confidence intervals for fatal crashes were curiously narrow. In fact, they were not very different from those of all injury crashes, which were more than 40 times as numerous. As the numbers used to calculate the odds ratio were not given, it is not possible to check the analyses, but simulations using estimated numbers (from percent ESC in the sample, raw numbers, and the odds ratios in this study) indicate much wider limits. Similarly, in Green and Woodrooffe (2006a), some of the confidence intervals did not match the raw data, although the effect sizes did.

The study by Koisaari, Kari, Vahlberg, Sihvola, and Tervo (2019) featured a very different methodology as compared to all others, as they only included culpable crash involvements in their study group. As could be expected (af Wåhlberg, 2008; 2009a; 2018), the effects in this paper were substantially larger than for others, with seven out of eight values being the highest ever reported for its category. This does not mean that the effects are considered erroneous, but that they are not fully comparable to the others.

Similarly, the study by Lyckegaard, Hels, and Bernhoft (2015) reported very high effects, especially given the fact that these authors used all other crashes as controls for singles. As for example multi-vehicle crashes have been found to show reductions too, this method should under-estimate, rather than over-estimate, the effects. However, the authors noted that a test using rear-end crashes as control did not yield any significant differences. These results can therefore be said to be an anomaly, which showed up as a double outlier (Ben-Gali, 2005) in several moderator analyses in the present study. Again, this does not necessarily indicate an error, but rather some unknown difference in methodology versus other studies.

3.2. First analysis; descriptives

In Table 5, unweighted averages and other descriptive statistics for the data gathered can be seen. Although there were 23 different studies, 57 different samples and 172 effects in this dataset, the maximum number of effects available in a variable (crash type) was 22, which is a first indication of the differences of the literature (different papers have used different crash types). Effect size data were entered into 18 variables (see Table 5).

There were 12 samples for LTV, 18 for passenger cars, and 27 mixed. Forty-one samples used matched models and 16 all kinds of vehicles. Twenty-six samples had used rear-end crashes as the control, 28 preferred other types, and 3 used exposure measures such as registered vehicles. Other descriptive results are displayed in Table 4 and 6. It is noteworthy that the percent of ESC in different samples differ very strongly, as do the sample sizes. It is necessary to repeat that these sample sizes are only for the totals, while effects in most cases were calculated for sub-samples, but these numbers were rarely reported.

3.3. Second analysis; separate for each crash variable

This part of the analysis aimed to calculate population estimates of effects for different crash categories, investigate possible publication bias and heterogeneity, and tentatively assess the influence of some moderators.

Table 7 displays the results for the variables that had at least eight data points (k) with confidence intervals for the odds ratios. Method effects were significant only in a few instances. In several cases, no effects could be calculated, because all samples were of the same type.

For six out of seven variables, significant amounts of heterogeneity were detected in the Q test (Huedo-Medina, Sanchez-Meca, Marin-Martinez, & Botella, 2006), despite k being rather small in all instances. This result was in agreement with the excess variance estimated by the I^2 method.

To test whether restricting the definition of ESC-relevant crashes lead to larger effects, the percent of case crashes of the total was calculated and correlated with the effects for all crash categories with more than 10 values on case crashes (see Table 7). These calculations were strongly dominated by the results of Dang (2007), one of the few to report numbers of cases for each sub-sample/analysis. In line with the hypothesis, all correlations were positive, indicating that a lower percent of cases (indicating a more restricted definition of ESC-relevant crashes) was associated with larger effects.

3.4. Third analysis; average of sample effects and moderators

As the second analysis suffered from the problem of results being scattered over many different outcome variables and the resulting low statistical power to detect any moderator effects, a different type of analysis was undertaken in two variants, by averaging all effects within each sample.

In the first method, the effects within each row (sample) of data were averaged, this effect was turned into an odds ratio, and finally these values were converted into Pearson r values (Poom & af Wåhlberg, 2022). The advantage of the last conversion is that r values can be *meta*-analyzed with sample size as a measure of reliability, which is not possible to do with odds ratios in the CMA software. It should be remembered that the sample sizes used were for the total samples, as the sub-sample data were usually not available. However, as the effects had been averaged from all parts of the sample, the total number should be most relevant for these effects. This dataset was called the averaged dataset.

In the second method, to remove variance due to crash categories having different population values, each effect size variable (e.g., fatal singles) was standardized against the overall mean of the variable (and therefore called the standardized dataset), whereafter these values were averaged for each sample. This dataset was analyzed with standard statistical methods, as the effect variable was not expressed in a standard effect size unit.

These methods yielded two variables with 57 effects each, which could be tested for moderator influences. In a *meta*-regression analysis of the averaged dataset (Table 8), three moderator variables were significant. In the analysis of the standardized dataset, three variables yielded significant differences versus effects (Tables 9-10), but only two of these were the same as in the *meta*-regression. This could be due to differences between the statistical methods in terms of weighting, but also interactions between moderators. A stepwise regression analysis was therefore run on the standardized dataset. This yielded the same result as the correlation analysis in terms of significant predictors.

The analyses of the two larger datasets thus agreed that effects were smaller for passenger cars and mixed samples as compared to LTVs, and that samples with matched vehicles yielded larger effects than non-matched cars (see Table 11). There were also indications that several of the other moderators had effects, but these were not stable over analyses, or not significant in the total sample.

Finally, calculations were made on how much the moderators influenced the effects in terms of absolute values (see Table 12). This was only undertaken in the averaged dataset, as the standardized one did not contain absolute values.

4. Discussion

4.1. Results

The main result would seem to be that the data were extremely heterogeneous, with many outlying values and some statistically significant moderators. The mean effects calculated were not very different from those of previous *meta*-analyses, but these values were not the prime target of interest in this study, but the unexplained heterogeneity in the data.

It was predicted that effects would be reduced over time, due to self-selection of drivers being most noticeable at the start of market penetration. Only very weak support for this hypothesis was found. However, as market penetration was probably not very high in most of the countries and time periods where the present data were gathered (about 17% in the United States in 2010, the midpoint of the last data from that country), it is possible that this effect will emerge in future studies. Behavioral adaptation could not be separately tested from the self-selection problem, as they would yield the same effects, but evidence for these effects is very weak in these data.

The type of vehicle studied was found to yield reliably different effects, as many researchers have posited. However, mixed samples did not yield a larger effect than passenger cars, which is peculiar.

No reliable effects for induced exposure versus other methods could be found, as there were preciously few studies that had not used this methodology, and one of them (Koisaari et al., 2019) was very different from all other studies, as discussed above. Similarly, no differences were found for different crash types used as non-targets in the induced exposure studies.

An association between percent ESC in samples and their effect sizes was something that at times would be present in the tests of different models, exclusion of outliers, etc. The meaning of increasing effects with a higher percent ESC is not easy to interpret, however. It might be an effect of increased variance in the data or be due to the elusive crash category problem.

The reporting of various important statistics, especially the sample size in each effect calculation and information was sorely lacking in many papers. Using the total sample sizes, it can only be concluded it is possible that effects are smaller for larger samples.

Finally, matching vehicles yielded reliably stronger effects than non-matching.

4.2. Comparisons with previous reviews and meta-analyses

Many studies on ESC have been published since the *meta*-analyses by Erke (2008) and Høye (2011). The current analysis is therefore larger in terms of data than the previous ones (15 samples published after 2011). On the other hand, some studies included in those analyses were excluded, and some not found.

As noted in the section Studies and Data Excluded, several studies that have been cited as evidence concerning ESC were excluded from analysis in the present study. Those that were included in previous *meta*-analyses will be discussed here.

Høye (2011, Table 1) included a study by Padmanaban et al. (2008) in her *meta*-analysis of ESC effects, stating that it contained 18 effect estimates. However, in the reference given, there are no such estimates of crash risk, as the paper is about various antecedents to crashes only for cars with ESC.

In the present study, the study by Farmer (2006) was excluded due to overlap with Dang (2007), while Høye (2011) included both. On the other hand, neither Høye nor Erke included Bahouth (2006). Most of those results were the same as Bahouth (2005), which was included by both authors.

The weights given by Høye for different samples are difficult to understand. For example, the statistical weight given for Farmer (2006) was some 50 times higher than that for Page and Cuny (2006), if these are interpreted as the sum for each effect and considering the difference in sample sizes. Furthermore, inconsistencies between the weightings applied by Erke and Høye would seem to be apparent, as the values for Aga and Okada (2003) and Page and Cuny (2006) are the same, but for Kreiss, Schüler and Langwieder (2006) it was different. As these authors included different numbers of estimates from the same studies, it is not possible to compare with the other values. No explanations for these phenomena have been found.

Neither Erke (2008) nor Høye (2011) described how they could apply statistical weights to samples that lacked this information. It is possible that they used the total sample sizes instead of sub-samples, but this would mean that their confidence intervals were much smaller than warranted.

In terms of average effects, no large differences to Erke (2008) and Høye (2011) were found in the present study. The most important difference to previous analyses of ESC, however, is the suggestion of previously unknown moderators in the data, and the analyses to find these.

4.3. Limitations

One considerable problem involved in *meta*-analysis of effects of crash prevention strategies is that different studies define their target populations differently, and therefore yield different results. Authors who define their targets very narrowly will probably find larger effects, because the intervention is more relevant to their cases (see for example Riexinger et al., 2019). This is a result of the lack of standard definitions of crashes in different countries and within research, and agreement about how evaluation studies of traffic safety should be undertaken. For the *meta*-analyst, this means greater heterogeneity in the data, and this may suppress important effects. In the present study this means that the suggested moderator of inflated effect sizes due to restricted definitions of crashes was very difficult to estimate with the available data. This kind of problem occurred for several of the moderators and is the strongest shortcoming of this study; the hypotheses have not been thoroughly tested.

One possible solution for the problem of different crash categories would be to re-calculate each effect into the effect on crashes in general, or for a certain severity level (as done by Chouinard & Lécuyer, 2011; Dang, 2007). This would enable the pooling of all results into a single variable, with much more statistical power and reliability of the findings. This method, however, was outside the scope of the present paper.

Another problem that impacts on the validity of the findings is that several authors have published different versions of the same data as well as updates of the same data (notably Lie, Tingvall, Krafft, & Kullgren, 2004; Lie et al., 2005; 2006; and probably Tingvall et al., 2022 referenced in Burton, Delaney, Newstead, Logan, & Fildes, 2004), as can be seen in Tables 2 and 3. Also, at times, different authors have used the same data. This double, triple, or quadruple publication on the same or overlapping datasets has not always been acknowledged by reviewers and other users of publications on ESC (e.g., Høye, 2011; Mackenzie, 2015), which could lead to an impression of there being more evidence available than is the case. In the present *meta*-analysis, several such instances were identified, but total certainty is difficult to achieve in this respect. Some of the results analyzed here may therefore not be fully independent.

As pointed out by an anonymous reviewer, this paper has not addressed a host of other possible moderators of effects, such as driver age, optional fitment of ESC, vehicle age effects, and changes in the environment. It is acknowledged that such factors might be one explanation for the heterogeneity of the data. However, as few studies included any measurements of any of these factors, it was not possible to test these as moderators. To elaborate a bit on these factors, it could be expected that driver (owner) age would differ between ESC and non-ESC cars, as older people tend to be more safety-conscious (i.e., this would be part of the suspected self-selection effect). For many brands of cars, there has been a period of time when certain models have been offered with ESC as an optional extra. Such in-between groups have often been excluded from analysis, because data on whether ESC was actually fitted have not been available. Vehicles with ESC as standard equipment would usually be a few years younger than those without, and additional safety features, such as improved crashworthiness might have been added and increased the difference between ESC and non-ESC. Finally, changes in the driving environment might have been made that have favored ESC-equipped cars, although such a mechanism would need to be rather complicated.

5. Conclusions

The most impelling result of this *meta*-analytic study concerns the heterogeneity of the available effects as well as the methods used to estimate them. Also, information that is needed for more thorough tests of the hypotheses about systematic over-estimations of effects presented here is lacking from most studies (mainly sub-sample sizes).

However, the current analysis and interpretations are speculative, and should be shored up with further research, using different methods to estimate the effect of ESC. This would include to avoid the self-selection problem, for example by studying the effect in a company fleet of cars, using actual exposure data and analyzing the effect over time. A totally different method would be to test whether the percent of ESC-relevant crashes has declined in the population when ESC has become more prevalent. Such a study has been undertaken on U.S. fatal crash data (af Wåhlberg & Dorn, 2024), and shows much smaller effects than the ones calculated here. This kind of study should be repeated in other countries with good crash records and high prevalence of ESC, such as several EU members. In general, most western countries seem to have made ESC mandatory equipment for new passenger vehicles before 2015 or so, and market penetration rates were probably well above 80% in 2020 (Bansal & Kockelman, 2017; Baum, Grawenhoff, & Geißler, 2007, Table 16; Krafft, Kullgren, Lie, & Tingvall, 2009; Weekes, Thatcham, Frampton, & Thomas, 2009).

In general, when new safety technology is tested, self-selection must be avoided, and some sort of experimental method be applied, with random allocation of subjects as the most important part. Furthermore, the groups must be studied over fairly long time periods (years), using both behavioral and crash data. Only when the technology is highly prevalent in the population is it possible to use field data to study the trend over time. Comparing vehicles with and without the technology without experimental control, however, will never be feasible, as self-selection will always be a problem.

The current paper concentrated upon the problem of the effect of ESC for light four-wheel vehicles, thus excluding trucks. It is predicted that estimates of effects of ESC for trucks will be smaller, as the effect studies on such vehicles have usually used exposure data from transportation companies. Also, self-selection at the driver level should be less of a problem with this population, although safer companies might be more prone to buying safety equipment.

Although the current results in general indicate sizeable effects for certain crash categories (>30%), the exact value is not possible to pinpoint with any acceptable accuracy, because important information is missing from the literature. More importantly though, the methodology used in ESC studies yields results that are too heterogeneous to be due to random error, and the main methodological problems identified in the present paper are difficult to test.

The current results therefore call into question the conclusions from all kinds of studies that have calculated or discussed expected benefits of ESC at the national level (e.g., Antona-Makoshi et al., 2023; Baum, Grawenhoff, & Geißler, 2007; Fitzharris, Scully, & Newstead, 2010; Fisa, Musukuma, Sampa, Musonda, & Young, 2022; Flannagan & Flannagan, 2009; Lukianov, 2009; Moennich et al., 2019; NHTSA, 2011; Page, Cuny, Zangmeister, Kreiss, & Hermitte, 2009; Tingvall et al., 2022; Vaa, Penttinen, & Spyropoulou, 2007; Wang, Zhong, Ma, Abdel-Aty, & Park, 2020; Weekes, Thatcham, Frampton, & Thomas, 2009). As analyzed here, the input values of ESC effectiveness used are not very reliable. It is noteworthy that these studies are very numerous, far exceeding the number of empirical studies that have actually studied the problem. Such resources could be better spent on the kind of research described above.

The practical implications of this study can be seen in two different domains: research and policy. There would seem to be a need to use research methods that bypass the problems of self-selection and behavioral adaptation, or measure and control for these. Otherwise, we can expect the same situation to arise with the next generation of automated features on vehicles. As for policy, ESC has been mandatory on new cars for decades in many countries, a principle that might have led to other safety interventions not being implemented. If there is widespread belief, and resources, going into a technology that is not as efficient as believed, this will probably lead to sub-optimal results.

CRediT authorship contribution statement

A.E. af Wåhlberg: Conceptualization, Data curation, Formal analysis, Methodology, Writing – original draft. **L. Dorn:** Funding acquisition, Project administration, Writing – review & editing.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgement

This work is supported by the Engineering and Physical Sciences Research Council [grant number: EP/V026763/1].

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.jsr.2024.07.004.

References

Aga, M., & Okada, A. (2003). Analysis of Vehicle Stability Control (VSC)'s Effectiveness from Accident Data. Paper Number 541. In *Proceedings of the 18th International Technical Conference on the Enhanced Safety of Vehicles.* Nagoya: Japan.

Antona-Makoshi, J., Muslim, H., Medojevic, M., Watanabe, S., Seguí-Gómez, M., & Bhalla, K. (2023). Estimated potential death and disability averted with vehicle safety interventions, Association of Southeast Asian Nations. *Bulletin of the World Health Organization, 101*(3), 211–222.

Aschenbrenner, K. M., & Biehl, B. (1994). Improved safety through improved technical measures? In R. Trimpop and G. Wilde (Eds.) *Challenges to Accident Prevention. The Issue of Risk Compensation Behaviour,* pp. 81-89. Groningen: Styx.

Bahouth, G. (2005). Real world crash evaluation of Vehicle Stability Control (VSC) technology. *Association for the Advancement of Automotive Medicine 49th Annual Conference*, pp. 19-34.

Bahouth, G. (2006). Reductions in Crash Injury and Fatalities Due to Vehicle Stability Control Technology. *SAE Technical Paper No.* 2006–01-0925.

Bansal, P., & Kockelman, K. M. (2017). Forecasting Americans' long-term adoption of connected and autonomous vehicle technologies. *Transportation Research Part A: Policy and Practice, 95*, 49–63.

Baum, H., Grawenhoff, S., & Geißler, T. (2007). Cost-benefit analysis of the Electronic Stability Program (ESP). *Zeitschrift fur Verkehrswissenschaft, 78*(3), 192–223.

Bayly, M., Fildes, B. N., Regan, M., & Young, K. (2007). *Review of Crash Effectiveness of Intelligent Transport Systems.* Deliverable D4. 1.1–D6. 2. Project No 027763 TRACE.

Becker, L., Delaney, A., & Newstead, S. (2007). *Safer Vehicle Purchases: Developing Cost/Benefit Estimates for Fleet Managers and Others - Part A: Crash Profiles and Risk Estimates.* AustRoads: Draft Report.

Ben-Gali, I. (2005). Outlier detection. In O. Maimon, & L. Rockach (Eds.), *Data Mining and Knowledge Discovery Handbook: A Complete Guide for Practitioners and Researchers*. Kluwer Academic Publishers.

Blower, D. (2014). *Assessment of the Effectiveness of Advanced Collision Avoidance Technologies*. University of Michigan Transportation Research Institute. Report UMTRI-2014-3. Ann Arbor:.

Branas, C. C., & Knudson, M. M. (2001). Helmet laws and motorcycle rider death rates. *Accident Analysis & Prevention, 33*(5), 641–648.

Broughton, J., & Baughan, C. (2002). The effectiveness of antilock braking systems in reducing accidents in Great Britain. *Accident Analysis & Prevention, 34*(3), 347–355.

Burton, D., Delaney, A., Newstead, S., Logan, D., & Fildes, B. (2004). *Effectiveness of ABS and Vehicle Stability Control Systems*. Royal Automobile Club of Victoria. RACV Research Report 00/04/01.

Chouinard, A., & Lécuyer, J. F. (2011). A study of the effectiveness of Electronic Stability Control in Canada. *Accident Analysis & Prevention, 43*(1), 451–460.

Dang, J. N. (2004). *Preliminary Results Analyzing the Effectiveness of Electronic Stability Control (ESC) Systems Evaluation Note*. Washington, D.C.: National Highway Traffic Safety Administration. DOT HS 809 790.

Dang, J. N. (2007). Statistical Analysis of the Effectiveness of Electronic Stability Control (ESC) Systems: Final Report. NHTSA Technical Report DOT HS 810 794. Washington, DC: National Highway Traffic Safety Administration.

Dorn, L., & af Wåhlberg, A. E.. (2018). Behavioural culpability for traffic accidents. *Transportation Research Part F: Traffic Psychology and Behaviour, 60*, 505–514. https://doi.org/10.1016/j.trf.2018.11.004

Erke, A. (2008). Effects of electronic stability control (ESC) on accidents: A review of empirical evidence. *Accident Analysis and Prevention, 40*, 167–173.

Erke, A. (2009). Red light for red-light cameras? A meta-analysis of the effects of red-light cameras on crashes. *Accident Analysis and Prevention, 41*, 897–905.

Fach, M., & Ockel, D. (2009). Evaluation methods for the effectiveness of active safety systems with respect to real world accident analysis. In *International Technical Conference on the Enhanced Safety of Vehicles*. Stuttgart.

Farmer, C. M. (2001). New evidence concerning fatal crashes of passenger vehicles before and after adding antilock braking systems. *Accident Analysis and Prevention, 33*, 361–369.

Farmer, C. M. (2004). Effect of Electronic Stability Control on automobile crash risk. *Traffic Injury Prevention, 5*, 317–325.

Farmer, C. M. (2006). Effects of Electronic Stability Control: An update. *Traffic Injury Prevention, 7*, 319–324. https://doi.org/10.1080/15389580600846273

Farmer, C. M. (2010). *Effects of Electronic Stability Control on Fatal Crash Risk*. Arlington: Insurance Institute for Highway Safety.

Ferguson, S. A. (2007). The effectiveness of electronic stability control in reducing realworld crashes: A literature review. *Traffic Injury Prevention, 8*(4), 329–338.

Field, A. P., & Gillett, R. (2010). How to do a meta-analysis. *British Journal of Mathematical and Statistical Psychology, 63*, 665–694.

Fildes, B., Keall, M., Thomas, P., Parkkari, K., Pennisi, L., & Tingvall, C. (2013). Evaluation of the benefits of vehicle safety technology: The MUNDS study. *Accident Analysis & Prevention, 55*, 274–281.

Fisa, R., Musukuma, M., Sampa, M., Musonda, P., & Young, T. (2022). Effects of interventions for preventing road traffic crashes: An overview of systematic reviews. *BMC Public Health, 22*(1), 513.

Fitzharris, M. (2020). Electronic stability control and side impact crashes: 100% cure or a case of realigning safety priorities. *Safety, 2011*, 23–25.

Flannagan, C. A., & Flannagan, M. J. (2009). *New Developments in UTMOST: Application to Electronic Stability Control.* UMTRI-2009-15. Ann Arbor: University of Michigan.

Frampton, R., & Thomas, P. (2007). *Effectiveness of Electronic Stability Control Systems in Great Britain.* PPAD 9/33/99. Loughborough University.

German Insurance Association. 2014. *Accident Statistics and the Potential of Driver Assistance Systems.* Report 47. German Insurance Association, Berlin.

Girasek, D. C., & Taylor, B. (2010). An exploratory study of the relationship between socioeconomic status and motor vehicle safety features. *Traffic Injury Prevention, 11* (2), 151–155.

Green, P. E., & Woodrooffe, J. (2006a). *The Effectiveness of Electronic Stability Control on Motor Vehicle Crash Prevention.* UMTRI-2006-12. Ann Arbor: University of Michigan.

Green, P. E., & Woodrooffe, J. (2006). The estimated reduction in the odds of loss-of-control type crashes for sport utility vehicles equipped with electronic stability control. *Journal of Safety Research, 37*(5), 493–499.

Grömping, U., Weimann, U., & Menzler, S. (2004). Split register study: A new method for estimating the impact of rare exposure on population accident risk based on accident register data. *1st International ESAR Conference. Hannover Medical School*.

Haight, F. A. (1973). Induced exposure. *Accident Analysis and Prevention, 5*, 111–126.

Hedlund, J. (2000). Risky business: Safety regulations, risk compensation, and individual behavior. *Injury Prevention, 6*(2), 82–89.

Herms, B. F. (1972). Pedestrian crosswalk study: Accidents in painted and unpainted crosswalks. *Highway Research Record, 406*, 1–13.

Hertz, E., Hilton, J., & Johnson, D. (1998). *Analysis of the Crash Experiences of Vehicles Equipped with Antilock Braking Systems – An Update.* NHTSA Technical Report No. 808 758. Washington, DC: National Highway Traffic Safety Administration.

Higgins, J. P., & Thompson, S. G. (2002). Quantifying heterogeneity in meta-analysis. *Statistics in Medicine, 21*, 1539–1558.

Høye, A. (2011). The effects of electronic stability control (ESC) on crashes-An update. *Accident Analysis and Prevention, 43*(3), 1148–1159.

Huedo-Medina, T., Sanchez-Meca, J., Marin-Martinez, F., & Botella, J. (2006). Assessing heterogeneity in meta-analysis: Q statistic or I2 index? *Psychological Methods, 11*, 193–206.

Hunter, J. E., & Schmidt, F. L. (1990). *Methods of Meta-analysis: Correcting Error and Bias in Research Findings*. Newbury Park, CA: Sage.

Ioannidis, J. P., Munafò, M. R., Fusar-Poli, P., Nosek, B. A., & David, S. P. (2014). Publication and other reporting biases in cognitive sciences: Detection, prevalence, and prevention. *Trends in Cognitive Sciences, 18*, 235–241.

Jiang, X., Lyles, R. W., & Guo, R. (2014). A comprehensive review on the quasi-induced exposure technique. *Accident Analysis & Prevention, 65*, 36–46.

Johnson, N. S., & Gabler, H. C. (2015). Reduction in fatal longitudinal barrier crash rate due to electronic stability control. *Transportation Research Record, 2521*(1), 79–85.

Kahane, C. J. (2014). *Updated Estimates of Fatality Reduction by Electronic Stability Control. DOT*. Washington DC: National Highway Traffic Safety Administration.

Kahane, C., & Dang, J. (2009). *The Long Term Effect of ABS in Passenger Cars and LTVs.* Technical Report DOT HS 811 182, National Highway Traffic Safety Administration, Washington DC.

Kallan, M. J., & Jermakian, J. S. (2008). *SUV rollover in single vehicle crashes and the influence of ESC and SSF*. Association for the Advancement of Automotive Medicine.

Keall, M., & Newstead, S. (2009). Selection of comparison crash types for quasi-induced exposure risk estimation. *Traffic Injury Prevention, 10*(1), 23–29.

Keall, M., & Newstead, S. (2021). *VSRG Research Program: Evaluation of ESC and SUV Roll Stability Control.* Report No 357. Monash: Monash University Accident Research Centre.

Koisaari, T., Kari, T., Vahlberg, T., Sihvola, N., & Tervo, T. (2019). Crash risk of ESC-fitted passenger cars. *Traffic Injury Prevention, 20*(3), 325–331.

Koppel, S., Charlton, J., Fildes, B., & Fitzharris, M. (2008). How important is vehicle safety in the new vehicle purchase process? *Accident Analysis & Prevention, 40*(3), 994–1004.

Koppel, S., Clark, B., Hoareau, E., Charlton, J. L., & Newstead, S. V. (2013). How important is vehicle safety for older consumers in the vehicle purchase process? *Traffic Injury Prevention, 14*(6), 592–601.

Krafft, M., Kullgren, A., Lie, A., & Tingvall, C. (2009). From 15% to 90% ESC penetration in new cars in 48 months-the Swedish experience. In *21st International Conference on the Enhanced Safety of Vehicles (ESV), Stuttgart, Germany*.

Kreiss, J. P., Schüler, L., & Langwieder, K. (2005). The effectiveness of primary safety features in passenger cars in Germany. In *In Proceedings of the 19th ESV Conference. Paper No* (pp. 05–0145).

Lie, A. (2012). Nonconformities in real-world fatal crashes- Electronic Stability Control and Seat Belt Reminders. *Traffic Injury Prevention, 13*(3), 308–314. https://doi.org/10.1080/15389588.2011.653842

Lie, A., Tingvall, C., Krafft, M., & Kullgren, A. (2004). The effectiveness of ESP (electronic stability program) in reducing real life accidents. *Traffic Injury Prevention, 5*(1), 37–41.

Lie, A., Tingvall, C., Krafft, M., & Kullgren, A. (2005). The effectiveness of ESC (Electronic Stability Control) in reducing real life crashes and injuries. In *Proceedings: International Technical Conference on the Enhanced Safety of Vehicles, 2005*. National Highway Traffic Safety Administration.

Lie, A., Tingvall, C., Krafft, M., & Kullgren, A. (2006). The effectiveness of Electronic Stability Control (ESC) in reducing real life crashes and injuries. *Traffic Injury Prevention, 7*(1), 38–43. https://doi.org/10.1080/15389580500346838

Lukianov, G. (2009). An examination of crash scenarios and the potential benefits of six advanced safety technologies. In: Proceedings of the 21st International Technical Conference on the Enhanced Safety of Vehicles (ESV). Stuttgart, Germany.

Lyckegaard, A., Hels, T., & Bernhoft, I. M. (2015). Effectiveness of electronic stability control on single-vehicle accidents. *Traffic Injury Prevention, 16*(4), 380–386.

Lyles, R. W., Stamatiadis, P., & Lighthizer, D. R. (1991). Quasi-induced exposure revisited. *Accident Analysis and Prevention, 29*, 275–285.

Mackenzie, J. R. (2015). *Crash Avoidance by Electronic Stability Control on Australian High Speed Rural Roads: An Analysis of Braking Interventions*. University of Adelaide. Doctoral dissertation.

MacLennan, P. A., Marshall, T., Griffin, R., Purcell, M., McGwin, G., & Rue, L. W. (2008). Vehicle rollover risk and electronic stability control systems. *Injury Prevention, 14*(3), 154–158.

Moennich, J., Lich, T., Kumaresh, G., & Georgi, A. (2019). Loss of control car accidents on Indian Roads - Benefit estimation of ESC. SAE Technical Paper 2019-26-0009. doi:10.4271/2019-26-0009.

Nhtsa. (2011). *Estimating Lives Saved Annually by Electronic Stability Control*. Washington DC: National Highway Traffic Safety Administration. DOT HS 811 545.

Oh, M., Youn, S., Jeong, E., & Oh, C. (2017). Evaluating traffic safety benefits of Electronic Stability Control system using a meta analysis: Focused on accident rates. *Journal of Korean Society of Transportation, 35*(4), 307–320.

Padmanaban J. (2007). Field performance study of Electronic Stability Control system effectiveness in US fatal crashes. In *2007 International IRCOBI Conference on the Biomechanics of Impact,* p. 19-21.

Padmanaban, J., Shields, L. E., Scheibe, R. R., & Eyges, V. E. (2008). A comprehensive review of rollover accidents involving vehicles equipped with Electronic Stability Control (ESC) systems. In *Annals of Advances in Automotive Medicine/Annual Scientific Conference*, *52,* p. 9. Association for the Advancement of Automotive Medicine.

Page, Y., & Cuny, S. (2005). *Is ESP effective on French roads?* (p. 55) Unterreihe Fahrzeugtechnik: Berichte Der Bundesanstalt Fuer Strassenwesen.

Page, Y., & Cuny, S. (2006). Is electronic stability program effective on French roads? *Accident Analysis & Prevention, 38*(2), 357–364.

Page, Y., Cuny, S., Zangmeister, T., Kreiss, J.-P., & Hermitte T. (2009). The evaluation of the safety benefits of combined passive and on-board active safety applications. In: 53rd Annual Proceedings, Association for the Advancement of Automotive Medicine (AAAM) pp. 117-127. Baltimore, MD.

Peltzman, S. (1975). The effects of automobile safety regulation. *Journal of Political Economy, 83*(4), 677–725.

Poom, L., & af Wåhlberg, A. E.. (2022). Accuracy of conversion formulas for effect sizes: A Monte Carlo simulation. *Research Synthesis Methods, 13*(4), 508–519. https://doi.org/10.1002/jrsm.1560

Riexinger, L., Sherony, R., & Gabler, H. (2019). *Has Electronic Stability Control Reduced Rollover Crashes?* SAE Technical Paper 2019-01-1022. doi:10.4271/2019-01-1022.

Rudin-Brown, C., & Jamson, S. (2013). *Behavioural Adaptation and Road Safety: Theory.* Evidence and Action: CRC Press.

Rudin-Brown, C. M., Jenkins, R. W., Whitehead, T., & Burns, P. C. (2009). Could ESC (Electronic Stability Control) change the way we drive? *Traffic Injury Prevention, 10* (4), 340–347.

Rudin-Brown, C. M., & Noy, Y. I. (2002). Investigation of behavioral adaptation to lane departure warnings. *Transportation Research Record, 1803*(1), 30–37.

Rudin-Brown, C. M., & Parker, H. A. (2004). Behavioural adaptation to adaptive cruise control (ACC): Implications for preventive strategies. *Transportation Research Part F: Traffic Psychology and Behaviour, 7*(2), 59–76.

Rumar, K., Berggrund, U., Jernberg, P., & Ytterbom, U. (1976). Driver reaction to a technical safety measure - studded tyres. *Human Factors, 18*, 443–454.

Sagberg, F., Fosser, S., & Sätermo, I.-A. (1997). An investigation of behavioural adaptation to airbags and antilock brakes among taxi drivers. *Accident Analysis and Prevention, 29*, 293–302.

Scully, J., & Newstead, S. (2007). *Preliminary Evaluation of Electronic Stability Control Effectiveness in Australasia.* Report No. 271. Clayton: Monash University Accident Research Centre.

Scully, J., & Newstead, S. (2008). Evaluation of Electronic Stability Control effectiveness in Australasia. *Accident Analysis and Prevention, 40*(6), 2050–2057.

Scully, J., & Newstead, S. (2010). *Follow-up Evaluation of Electronic Stability Control Effectiveness in Australia.* Report No. 306. Monash University Accident Research Centre: Clayton.

Sivinski, R. (2011). *Crash Prevention Effectiveness of Light-Vehicle Electronic Stability Control: An Update of the 2007 NHTSA Evaluation.* Washington, DC: National Highway Traffic Safety Administration. NHTSA Report No. DOT HS-811-486.

Soni, S., Reddy, N., Tsapi, A., van Arem, B., & Farah, H. (2022). Behavioral adaptations of human drivers interacting with automated vehicles. *Transportation Research Part F: Traffic Psychology and Behaviour, 86*, 48–64.

Stamatiadis, N., & Deacon, J. A. (1997). Quasi-induced exposure: Methodology and insight. *Accident Analysis and Prevention, 27*, 443–459.

Strandroth, J., Rizzi, M., Olai, M., Lie, A., & Tingvall, C. (2012). The effects of studded tires on fatal crashes with passenger cars and the benefits of electronic stability control (ESC) in Swedish winter driving. *Accident Analysis and Prevention, 45*, 50–60.

Summala, H. (1997). Hierarchical model of behavioural adaptation and traffic accidents. In T. Rothengatter, & E. Carbonell Vaya (Eds.), *Traffic and Transport Psychology: Theory and Application* (pp. 41–52). Amsterdam: Pergamon.

Thomas, P. (2006a). Crash involvement risks of cars with electronic stability control systems in Great Britain. *International Journal of Vehicle Safety, 1*(4), 267–281. https://doi.org/10.1504/ijvs.2006.011229

Thomas, P. (2006b). *The accident reduction effectiveness of ESC equipped cars in Great Britain.* London. Citeseer: World Congress.

Thomas, P., & Frampton, R. (2007). Real world assessment of relative crash involvement rates of cars equipped with Electronic Stability Control. In *20th International Technical Conference on the Enhanced Safety of Vehicles,* Paper 07-0184. Lyon.

Thulin, H. (2007). *Monitoring of the Regulation Obliging Drivers To Give Way To Pedestrians On Zebra Crossings.* VTI Rapport 597. Linköping: Statens Väg- och Transportforskningsinstitut.

Tingvall, C., Michael, J., Larsson, P., Lie, A., Segui-Gomez, M., Wong, S. V., & Furas, A. (2022). Saving lives beyond 2020: The next steps. In *The Vision Zero Handbook: Theory, Technology and Management for a Zero Casualty Policy* (pp. 1–52). Cham: Springer International Publishing.

Vaa, T., Penttinen, M., & Spyropoulou, I. (2007). Intelligent transport systems and effects on road traffic accidents: State of the art. *IET Intelligent Transport Systems, 1*(2), 81–88.

Vadeby, A., Wiklund, M., & Forward, S. (2011). Car drivers' perceptions of electronic stability control (ESC) systems. *Accident Analysis & Prevention, 43*(3), 706–713.

Viechtbauer, W., & Cheung, M. W. (2010). Outlier and influence diagnostics for meta-analysis. *Research Synthesis Methods, 1*(2), 112–125. https://doi.org/10.1002/jrsm.11

Wang, L., Zhong, H., Ma, W., Abdel-Aty, M., & Park, J. (2020). How many crashes can connected vehicle and automated vehicle technologies prevent: A meta-analysis. *Accident Analysis & Prevention, 136*, Article 105299. https://doi.org/10.1016/j.aap.2019.105299

Weekes, A., Avery, M., Frampton, R., & Thomas, P. (2009). In *ESC standard fitment and failure to protect young drivers* (pp. 15–18). Germany: International Congress Center Stuttgart.

Wilde, G. J. (1982). Critical issues in risk homeostasis theory. *Risk Analysis, 2*(4), 249–258.

af Wåhlberg, A. E.. (2008). The relation of non-culpable traffic incidents to bus drivers' celeration behavior. *Journal of Safety Research, 39*, 41–46. https://doi.org/10.1016/j.jsr.2007.10.009

af Wåhlberg, A. E.. (2009a). *Driver Behaviour and Accident Research Methodology; Unresolved Problems.* Farnham: Ashgate. http://www.ashgate.com/isbn/9780754670766.

af Wåhlberg, A. E.. (2009b). Hourly changes in accident risk for bus drivers. *Journal of Risk Research, 12*, 187–197. https://doi.org/10.1080/13669870802488891

af Wåhlberg, A. E.. (2018). The effect of driver improvement interventions on crash involvement; has it been under-estimated? *Transportation Research Part F: Traffic Psychology and Behaviour, 54*, 349–356. https://doi.org/10.1016/j.trf.2018.02.027

af Wåhlberg, A. E., & Dorn, L. (2007). Culpable versus non-culpable traffic accidents; what is wrong with this picture? *Journal of Safety Research, 38,* 453-459. PMID: 17884432, DOI: 10.1016/j.jsr.2007.01.013.

af Wåhlberg, A. E. & Dorn, L.. (2019). Experience as a safety factor in bus driving; methodological considerations in a sample of bus drivers. *Safety, 5*, 37. https://doi.org/10.3390/safety5020037

af Wåhlberg, A. E., & Dorn, L.. (2024). The effects of Electronic Stability Control on fatal crash rates in the US. *Journal of Safety Research, 88*, 217–229. https://doi.org/10.1016/j.jsr.2023.11.008

Yannis, G., Papadimitriou, E., Dupont, E., & Martensen, H. (2010). Estimation of fatality and injury risk by means of in-depth fatal accident investigation data. *Traffic Injury Prevention, 11*(5), 492–502.

A.E. af Wåhlberg receieved his PhD in psychology at Uppsala University (Sweden) in 2006, and is currently a Research Fellow at Cranfield University (UK). His research interests include traffic safety, social science methodology, meta-analysis, human factors and social change.

L. Dorn is an Associate Professor of Driver Behaviour at Cranfield University. Her main research interests are individual differences in driver behaviour and research methodology in traffic safety. She has worked with the public and private sector in the design of interventions to improve driver safety for 25 years and is currently Co-Investigator on a UKRI/EPSRC funded project on behavioural adaptation to automated systems.