Transportation Research Part F: Traffic Psychology and Behaviour, Volume 54, April 2018, pp, 349-356
DOI: 10.1016/j.trf.2018.02.027

The effect of driver improvement interventions on crash involvement; has it been under-estimated?

A. E. af Wåhlberg
Cranfield University
Cranfield, Bedfordshire, MK43 0AL
UK
aw@empirica.pm

Abstract
The available evidence suggests that driver improvement interventions (with the aim to increase driver safety, most often by education or training) do not work. The average effect calculated in several meta-analyses is close to, and not always possible to distinguish from, zero, despite total samples sizes of several hundred thousand drivers. However, it is possible that all studies included in these meta-analyses have under-estimated the effect, due to a methodological error; all crashes have been used as dependent variable, instead of only those that the targeted drivers have caused. This error is expected to have considerably deflated the effect sizes, but it is not known how large this effect could be.
Using crash data for bus drivers in which culpability had been reliably established, a simple simulation was performed to determine the difference between using culpable and all crashes as an estimator of a safety effect. Using data for six years, calculations were made on single years. About ten percent of culpable crashes in each year were deleted to simulate a safety effect, where after the difference between the original and the simulated variable were calculated, using culpable only and all crashes in parallel. The effects using these two different kinds of datasets could then be compared and the under-estimation effect estimated.
Culpable crashes, as compared to all crashes, yielded larger differences in means between time periods, and smaller standard deviations. In between-subjects comparisons resulted in 15-30 percent larger effects for culpable crashes. Within-subjects calculations yielded larger but not as systematic effects.
The effect of driver improvement on crash involvement has been systematically under-estimated, as extremely few evaluation studies seem to have taken culpability for crashes into account. Therefore, new evaluations need to be undertaken, and/or old data re-analysed, to calculate a better estimate of the true effect of training and education in driving safety.

Key words: evaluation; driver education; driver training; culpability; intervention

Published by Elsevier. This is the Author Accepted Manuscript issued with: Creative Commons Attribution Non-Commercial No Derivatives License (CC:BY:NC:ND 4.0).
The final published version (version of record) is available online at DOI:10.1016/j.trf.2018.02.027. Please refer to any applicable publisher terms of use.

1. Introduction

One of the few problems that traffic safety researchers seem to agree upon is that driver improvement (which in this paper is taken to include all kinds of interventions where the aim is to reduce crash rates, usually in the form of training and education, for both problem and beginner drivers) does not have any proven safety effects (i.e. crash counts are not reduced), or that the effects are so miniscule as to be of no practical significance. Most of the many reviews and meta-analyses on this subject have reached this conclusion (Klein, 1966; Kaestner, 1968; McGuire & Kersh, 1969; Lund & Williams, 1985; Struckman-Johnson, Lund, Williams & Osborne, 1989; Mayhew, Simpson, Williams & Ferguson, 1998; Vernick et al., 1999; Christie, 2001; Mayhew & Simpson, 2002; Masten & Peck, 2004; Ker et al., 2005; Strathman, Kimpel & Leistner, 2007; Lonero, 2008; Roberts, Kwan & Cochrane Injuries Group Driver Education Reviewers, 2008; Peck, 2011; see Table 1 for some effect sizes), possibly with some caveat for possible effects of some specific sort of curriculum, and drink driving interventions (Wells-Parker, Bangert-Drowns, Mcmillen & Williams, 1995). It can also be pointed out that most of the interventions included in these analyses were aimed at problem drivers (repeat violators and crashers) or young drivers (who have a higher risk than older drivers). The possibilities for achieving and finding an effect have therefore largely been maximised, at least in terms of statistics. Meanwhile, organisations that deal with transportation and safety have continued to educate and train drivers as if nothing had happened.

One of the problems in the evaluation of effects of driver improvement in traffic safety is the low variance of the accident variable. This yields very low power, so even if there is indeed an effect of training, this will be very difficult to detect. Recently, Peck (2011) calculated that we would need tens of thousands of drivers to achieve reasonable power (80% chance of detecting a ten percent reduction in crashes). Still, the number of studies on safety effects and the total number of subjects are rather large (see Table 1 for an overview of meta-analyses on this subject). However, several researchers have pointed out that the quality of most studies on improvement courses is doubtful (Lund & Williams, 1985; Lonero & Mayhew, 2010; Wundersitz & Hutchinson, 2006; Peck, 2011), in terms of not being randomized, having small sample sizes, using variables with doubtful validity etc.

It is well known that studies with doubtful methodology tend to report large effects (e.g. Rongen, Robroek, van Lenthe & Burdorf, 2013; af Wåhlberg, Barraclough & Freeman, 2016), and the evidence regarding driver education and training might therefore be even less positive than what might be believed, as whatever few positive effects have been found may be artefacts. The outlook for evidence in favour of driver training and education would therefore seem to be bleak.

The conclusion that driver improvement does not have a practically useful effect is thus based upon a fair amount of data, but the quality of this data can be questioned. The aim of the present paper is to present yet another problem inherent in evaluations of safety interventions, which might have systematically deflated the effect sizes in such studies.

First, it can be noted that a peculiar effect exists within driver improvement evaluation research; calculations on crashes and offences tend to yield different results (Struckman-Johnson, Lund, Williams & Osborne, 1989), with the latter having larger effects. In the Masten and Peck meta-analysis (2004), the difference was 100 percent for the effect sizes (d .03 for crashes versus 0.06 for violations), which is the same as in Ker, Roberts, Collier, Beyer, Bunn and Frost (2005). Expressed in percent, the difference was smaller (28%) in the Masten and Peck study (-6.49% versus -8.28%), while Ker et al., did not report the results in percent. This effect could possibly be explained by differences in variance, although an actual differential effect on crashes and violations is possible. Offences and crashes correlate only about .18 (Barraclough, af Wåhlberg, Freeman, Watson & Watson, 2016), and and it is

therefore very apparent that they to a large degree are due to different behaviours. Here, however, it will be suggested that part of the effect difference is due to a lack of validity of the crash variable.

Turning to how improvement evaluations are usually undertaken, there exists one methodological problem which has not been discussed in the literature, and which would impact negatively upon the effects found. This concerns the culpability for the crashes used in the evaluations. Usually, this is not taken into account, i.e. all crashes are used as dependent variable.

This is problematic, because driver training and education aim to change the accident-causing behaviour of the driver. It does not aim to reduce the exposure to risk from being hit by other drivers by no fault of their own. Therefore, if all crashes are used as the dependent variable, this variable will contain a fair amount of error variance, which reflects amount of exposure and not the quality of driving behaviour.

It has been proposed that in studies of individual differences in driving safety, only culpable crashes should be used as the dependent variable (af Wåhlberg, 2003; 2009), as this will yield a purer measure of the intended construct, and therefore larger effect sizes (as shown in af Wåhlberg, 2008; 2009). This logic can be extended to evaluations, as the basic problems are the same. If the dependent variable is a mixture of the intended construct and error data, differences will be smaller, and standard deviations larger, and the effect sizes therefore smaller.

Then remains the problem of establishing culpability for crashes. Without a valid method for such categorization, the concept of using only culpable events would be meaningless, which is probably the reason why most researchers use all accidents (af Wåhlberg, 2003; 2009). Two different methods for establishing culpability have been proposed, that of Robertson and Drummer (1994), and that of af Wåhlberg and Dorn (2007; see also Dorn & af Wåhlberg, submitted; af Wåhlberg, 2009). These two methods have similarities but also important differences. Both propose that culpable crashes will be the best measure of driver safety behaviour and test the validity of the categorization by comparing effect sizes. The differences are that while Robertson and Drummer constructed a graded quantitative scoring scheme specifically for use in impaired driving research, af Wåhlberg and Dorn relied on a dichotomous approach for scoring (culpable/not culpable) and applied this to all kinds of driver/accident research. Most importantly, however, the validity test of the latter method is more advanced, as it relies on a testable theory which yields a fair number of ways of testing the accuracy of the scoring.

The culpability categorization test proposed in af Wåhlberg and Dorn (2007; af Wåhlberg, 2009) uses the basic of assumption of induced exposure; that drivers not responsible for their crashes are a random sample from the driving population. This assumption is then used as a definition. If culpability has been correctly assigned in a sample, the drivers with non-culpable crashes should be a random sample of the population, and therefore be similar to the population on all kinds of variables. If not, the assignment of culpability has been incorrectly undertaken, and the criteria (similar to Robertson and Drummers scoring scheme, but not as explicitly defined) will have to be revised. The advantage of this method is that several tests can be undertaken, if suitable data is available. For example, driving experience should be negatively correlated to number of culpable crashes (in the last few years), but not to non-culpable ones.

Table 1: Effect sizes for some driver interventions versus crashes in various meta-analyses. A positive d means a reduction in crashes.

Study	Type of population	Type of intervention	Dependent variable	N educated/control or k	Effect size (confidence intervals), type of statistic
Ker et al., 2005	Average drivers	Remedial education	Crashes	180 563/54 380	0.98 (0.96-1.01) Risk ratio
Ker et al., 2005	Average drivers	Advanced education	Crashes	35 539/12 222	0.99 (0.93-1.05) Risk ratio
Ker et al., 2005	Average drivers	Remedial education	Injury crashes	34 186/23 263	1.17 (0.89-1.54) Risk ratio
Ker et al., 2005	Average drivers	Advanced education	Injury crashes	1103/537	0.94 (0.74-1.20) Risk ratio
Masten & Peck, 2004	Problem drivers	Education, sanctions etc	Crashes	Total 1 640 000	0.030 (0.027-0.034) Cohen's d -6.49 %
Roberts, Kwan & Cochrane Injuries Group Driver Education Reviewers, 2008	Pre-license drivers	Education	Crashes	Total 17 965	1.04 (1.02-1.05) Risk Ratio 1.01 (0.83-1.23) Risk Ratio 1.03 (0.98-1.09) Risk Ratio
Wells-Parker, Bangert-Drowns, Mcmillen & Williams, 1995	Drink/drive offenders	Education/counselling	Non-alcohol related Alcohol-related	k=5 k=10	-0.11 Cohen's d? 0.07 Cohen's d?

This random sample method was tested in two samples of bus drivers (from Sweden and the UK; af Wåhlberg & Dorn, 2007), with different percentages of culpable involvements. It was found that the Swedish data, with about 70 percent culpable incidents, was very close to the assumed characteristics of correctly assigned culpability. The UK data, on the other hand, with 50 percent culpable incidents, showed features of having many culpable incidents classified as non-culpable (for example a correlation between these categories). The same kind of results has also been shown in a different population of bus drivers from the UK (Dorn & af Wåhlberg, submitted).

Returning to the problem of driver education efforts being erroneously evaluated due to the use of all accidents as dependent variable, it can be noted that hardly any of the reviews and meta-analyses cited above mentioned culpability (at fault, responsibility) in their inclusion criteria or discussions (Wells-Parker, Bangert-Drowns, Mcmillen & Williams, 1995; Mayhew, Simpson, Williams & Ferguson, 1998; Vernick, et al., 1999; Christie, 2001; Mayhew & Simpson, 2002; Masten & Peck, 2004; Ker, Roberts, Collier, Beyer, Bunn & Frost, 2005; Strathman, Kimpel & Leistner, 2007; Roberts, Kwan & Cochrane Injuries Group Driver Education Reviewers, 2008; Peck, 2011). Some reviewers have noted that not all crashes are due to the behaviour of the drivers who have been trained (Struckman-Johnson, Lund, Williams & Osborne, 1989; Lonero & Mayhew, 2010), but this has not led to any methodological difference in their work (e.g. including a moderator variable for type of dependent variable).

It would seem that only a few driver improvement evaluation papers have reported the effect for culpable crashes only, or in addition to all crashes (Harrington, 1972; Kaestner, 1980; Musicant, Lotan & Toledo, 2007; af Wåhlberg, 2007). In the Harrington paper, the effect on this variable was larger than for all crashes, as expected, but this result should be treated with caution, as the percent at fault crashes (as compared to all) was very small (<20%). A large number of truly culpable crashes were therefore probably excluded from this variable, with effects that currently cannot be predicted. Kaestner (1980, only abstract available) reported a positive effect for 'chargeable accidents' and moving violations, so it is not possible to know which variable was carrying the effect (af Wåhlberg, 2009). In af Wåhlberg (2007), the results were very similar between variables, but as the effect of training on behaviour was very small (≈2% over one year), this could be expected. In Musicant, Lotan and Toledo (2007), the effect was larger for all collisions as compared to at fault (44/38%). However, only fifty-three percent of crashes were deemed culpable, apparently judged by an insurance company. Also, the medium risk group (grouped according to driving behaviour data gathered by a telematics system) had a lower percent of at fault crashes, which should not be the case. If exposure is similar, the rate of culpable crashes should increase as the absolute number of crashes per driver increases.

In the meta-analysis by Wells-Parker, Bangert-Drowns, Mcmillen and Williams (1995), crashes which involved alcohol were separated from the rest, and while the former were reduced, the latter increased slightly. This result can be interpreted in terms of alcohol-related crashes probably being due to the behaviour of the driver in the intervention study, while non-alcohol-related crashes would also include non-culpable events. This interpretation is very tentative, however, as there are some other possible effects involved which could influence the effects, such as drunk drivers who have received treatment driving more, and causing crashes, when they are sober but hung-over.

The empirical results would seem to be conflicting. However, the explanation for this probably lies in how culpability has been assigned. If a large number of actual at fault crashes are excluded, the effect will of course be reduced.

It can be therefore be suggested that future evaluations (or re-analyses of previous ones) use either of these responsibility methods to remove non-culpable accidents from the data before

doing comparisons of before and after intervention. However, it would be preferable to know something about what kind of a difference could be expected beforehand. This would allow researchers to estimate whether it would actually be worth the effort to do further research on this problem, but it would also be possible to re-calculate previous meta-analytic results to provide a more accurate population estimate for the effects of driver safety education and training. The more empirically oriented part of this paper was therefore aimed at trying to estimate the difference in effect when using all accidents and culpable accidents only in safety evaluations, using a simple simulation.

2. Method

2.1 Data

Bus accident data from the bus company Gamla Uppsalabuss (Uppsala, Sweden) was used. This bus company owns about 180 buses, employs about 350 drivers and run all intra-city bus routes in Uppsala, a town of some 200 000 inhabitants. The incidents had been coded in several different variables, including culpability for the event (tested in af Wåhlberg & Dorn, 2007). This data had been gathered in several projects and analysed in many different ways (af Wåhlberg, 2002; 2004), especially as a dependent variable in studies on driving behaviour (e.g. af Wåhlberg, 2006). Data for the period 2000-2005 was used, separated into single years (see Table 2 for descriptive statistics for one year).
The accidents which were coded as not culpable tended to be those where the bus driver had been shunted from being while being at a standstill, when the other driver committed a right of way violation, and when the other vehicle was out of control (i.e. skidding).

2.2 Simulation of a safety effect

First, two accident variables were constructed; all crashes and culpable only (per driver), for each year. The culpable crashes had been judged by the author, and the coding tested for correctness in af Wåhlberg and Dorn (2007; see also Dorn & af Wåhlberg, submitted). Thereafter, an intervention effect was simulated by randomly removing about ten percent of the culpable events in each year (i.e. the same incidents were removed from both crash variables). The effects of using culpable instead of all crashes could then be calculated by using the original variable as the control group, and the simulated data as the intervention group, for both the All crashes and the culpable only variable, computing the effect of the simulation. Finally, the effects could be compared between All and culpable only.
The described method simulates a situation where the effect is calculated between an intervention group and a control, in the same year. However, it is also possible to compare an intervention group over time. To simulate this situation, the data described were compared over time instead, i.e. the All accidents 2000 variable was compared against the simulated effect All accidents 2001, where after the effect for culpable accidents was computed in the same way, and the effect finally compared.

Table 2: Descriptive data for the bus driver sample in 2001 (calculated for the last day of the year). Shown are the percent of men, the means and standard deviations for experience, age (in years), and the hours worked in 2001. N=413.

Sex	Age	Experience (N=408)	Hours worked 2001
89.8%	45.5/10.3	10.4/10.4	1219/603

2.3 Analysis

Effects of interventions can be calculated in three different ways, two of which will be tested here. Commonly, an effect is computed between the receivers and a control group (between-subjects). However, it is also possible to compare the intervention group before and after the

intervention (within-subjects). Here, both these variants will be tested. The third will be treated in the discussion, as it is rather rare.
The difference between an intervention group and a control can be expressed statistically in many ways. It is very possible that the difference in effects analysed here differs somewhat between different ways of calculating the basic intervention effect. As to produce an estimate which is more flexible and comprehensive than a single statistic, it was decided to use several statistical methods in parallel.
For practical purposes, a percent value is probably most useful, as it can be entered into calculations on the economical effect of safety interventions. In scientific statistics, on the other hand, different types of effect measures which are based upon variance are usually used. For ease of comparisons with, and computations on, meta-analytic data, Cohen's d (similar to Hedge's g) would seem to be preferable, as this is commonly used in meta-analysis. As third and fourth metrics, the common t and ANOVA F values were chosen, as they have at times been used in the safety evaluation literature.

3. Results

In Table 2 can be seen the descriptive data for the bus drivers working in 2001, which overlap strongly with the other years. These values are very stable over time.

Table 3: The differences in effects between all accident and culpable accidents, given the same reduction in crashes, comparing the same year with and without an intervention effect. Mean 1 is for the real data, while Mean 2 is for the simulated data (culpable accidents removed). The differences between these two are expressed in ANOVA F and independent t-test values, Cohen's d and percent.

N	388		413		396	
	All crashes 2000	Culpable crashes 2000	All crashes 2001	Culpable crashes 2001	All crashes 2002	Culpable crashes 2002
Mean (std) 1	0.263 (0.564)	0.211 (0.484)	0.305 (0.622)	0.223 (0.534)	0.217 (0.506)	0.177 (0.471)
Mean (std) 2	0.242 (0.536)	0.191 (0.449)	0.281 (0.594)	0.199 (0.502)	0.199 (0.475)	0.159 (0.436)
F	0.28	0.38	0.33	0.45	0.26	0.30
t	0.52	0.62	0.57	0.67	0.51	0.55
d	0.038	0.043	0.040	0.047	0.036	0.039
Percent	-8.0%	-9.5%	-7.9%	-10.8%	-8.1%	-10.0%
N	419		387		355	
	All crashes 2003	Culpable crashes 2003	All crashes 2004	Culpable crashes 2004	All crashes 2005	Culpable crashes 2005
Mean (std) 1	0.279 (0.528)	0.220 (0.474)	0.315 (0.579)	0.235 (0.503)	0.318 (0.636)	0.239 (0.522)
Mean (std) 2	0.253 (0.506)	0.193 (0.477)	0.284 (0.555)	0.204 (0.469)	0.282 (0.601)	0.203 (0.480)
F	0.54	0.68	0.58	0.79	0.62	0.95
t	0.73	0.83	0.76	0.89	0.79	0.97
d	0.051	0.057	0.055	0.064	0.059	0.073
Percent	-9.4%	-12.0%	-9.8%	-13.2%	11.5%	15.3%

Table 3 show the real accident data as Mean 1, and the simulated effect of improvement as Mean 2 (Mean 2=Mean 1 with about ten percent of culpable crashes removed), and the differences between these calculated in different ways. It can be seen that the standard deviations were smaller for culpable crashes, and that the effects were always larger for this variable. In Table 4, these differences have been calculated as percent increase in effect when using the culpable crashes only as dependent variable. Correlating these values with the original improvement effect in percent showed that there was a strong positive association between these, i.e. the stronger the actual effect, the larger the difference between All and culpable crashes calculations. This means that larger effects have been more strongly under-estimated. For example, if the reduction in culpable crashes was ten percent, the difference

between culpable and All crashes for Cohen's d would be about twelve percent, but if the reduction was fifteen percent, the difference would be twenty-one percent.

Table 4: The differences between using culpable and all crashes as dependent variable for the different effect size measures, from the data in Table 3.

Year	Percent difference F	Percent difference t	Percent difference d	Percent difference percent
2000	35.7%	19.2%	13.2%	18.7%
2001	36.4%	17.5%	17.5%	36.7%
2002	15.4%	7.8%	8.3%	23.5%
2003	25.9%	13.7%	11.8%	27.7%
2004	36.2%	17.1%	16.3%	34.7%
2005	53.2%	22.8%	23.7%	33.0%
Mean	**33.8%**	**16.4%**	**15.1%**	**29.0%**

Thereafter, within-subjects effects were calculated by comparing the means in one year with the simulated effect in the next year (i.e. All crashes in 2000 compared with the simulated crash data in 2001), using repeated measurements ANOVA and dependent t-tests. In Tables 5 and 6 it can be seen that the trend for these calculations is similar, and on average even stronger than for the between-subjects results. However, these results are not as unanimous as the previous ones, as in one case the All crashes variable yields the stronger effect, and there is also a case where the effect for both types of crashes is perverse.

Table 5 The differences in effects between all accident and culpable accidents, given the same reduction in crashes, comparing different years. The differences are expressed in (repeated measures) ANOVA F and (dependent) t-test values, Cohen's d and percent.

N	342		371		337		374		333	
Years	2000/2001		2001/2002		2002/2003		2003/2004		2004/2005	
Accident variable	All crashes	Culpable crashes	All crashes	Culpable crashes	All crashes	Culpable crashes	All crashes	Culpable crashes	All crashes	Culpable crashes
F	1.97	5.39	14.14	7.35	1.05	0.38	0.08	0.52	1.52	1.71
t	1.40	2.32	3.76	2.71	-1.02	-0.62	0.27	0.72	1.23	1.31
d	0.11	0.18	0.25	0.19	-0.07	-0.04	0.02	0.05	0.09	0.10
Percent	20.8%	35.5%	41.9%	37.4%	-16.5%	-11.1%	3.6%	10.5%	16.1%	19.5%

Table 6: The mean of differences between using culpable and all crashes as dependent variable for the different effect size measures, from the data in Table 5.

Year	Percent difference F	Percent difference t	Percent difference d	Percent difference percent
2000/2001	173.6%	65.7%	63.6%	70.7%
2001/2002	-48.0%	-27.9%	-24.0%	-10.7%
2002/2003	63.8%	39.2%	42.9%	32.7%
2003/2004	550.0%	166.7%	150.0%	191.7%
2004/2005	12.5%	6.5%	11.1%	21.1%
Mean	**150.4%**	**50.0%**	**48.7%**	**61.1%**

4. Discussion

4.1 Results

The present paper has pointed out a basic error in almost all known meta-analyses and reviews of evaluations of driver improvement interventions which have used crashes as a dependent variable. This error is probably also present in the papers which the reviews are based upon, or this would have been noticed. This error has led to a mis-estimation of the actual effect, deflating it about 10-20 percent, especially when the true effect has been large. Also, when within-subjects tests are used, the average under-estimation is even larger,

although the variation between single estimates is also much larger than for between-subjects computations.

The present results were not dependent upon type of statistical method used, although the size of the error did differ. However, it is possible that larger simulations would reduce this difference, although it could of course be due to differences in the actual calculations of the statistical values.

One interesting feature of the present suggestion of a basic methodological error in all studies on driver education and training is that if it is true, it would explain why some studies have found that driver programs tend to reduce the count of offences, but not those of accidents (Lund & Williams, 1985). There are even a few cases where the accident count increased (Janke, 1994). This could partly be an effect of the offence data being purer than the accident data (mixed culpable and non-culpable crashes), although the rival explanation of drivers learning to avoid violating better than avoiding crashes is not ruled out. These two explanations can instead co-exist, and cause the rather large difference between effects for crashes and violations.

Furthermore, evaluations of driver improvement interventions have sometimes used intermediate (proxy) parameters as dependent variables, instead of crashes and violations. Although this question has not been studied, it could be suspected that these different variables have rather deviating effect sizes, in case this would need some sort of explanation. Several possibilities exist, such as social desirability, but this has not been systematically related to effect size differences.

It can be noted that the present results are applicable to all kinds of evaluations of safety-interventions where crashes, damage, costs or injuries are used as the dependent variable, unless the sample used is also the total population involved (as would be the case for injuries in a factory, if all workers were included). Therefore, the effect of Graduated Driving License schemes may also have been under-estimated, as they apparently do not take culpability into account (e.g. Zhu, Cummings, Chu, Coben & Li, 2013).

The calculations reported here have only considered the situation when there is an actual reduction in number of culpable crashes, and estimated how large this effect may be. However, as there also exist some reports of perverse effects, where driver education has increased the total number of crashes (in some cases due to earlier licensing; Roberts, Kwan & Cochrane Injuries Group Driver Education Reviewers, 2008), this situation should also be discussed. Simulating a perverse effect in the same way as the reductions presented here yields similar differences. Therefore, if a report has found a perverse effect on crashes, this is probably under-estimated too. In effect, the culpability method will tend to increase the effects found, whether positive or negative, with increasing strength as the original effect increases, thus making it easier to distinguish between the effects of different types of interventions.

Given that the effect sizes reported for driver improvement versus crashes in several meta-analyses are very close to zero, it might be argued that it does not matter much if the effect is actually some ten or even twenty percent larger. The usefulness of driver improvement programs would still be extremely limited (Helander, 1984, thought that the minimum effect size of practical importance would be a 10 percent reduction). However, many evaluations have also been criticised for other methodological shortcomings, as noted in the introduction. Little is known about whether these shortcomings have lead to under- or over-estimations of the true effect. Masten and Peck (2004) found that good quality studies had larger effects. Ker et al. (2005) coded for quality of the studies, but concluded that they could not perform a moderator analysis, as not enough information was available. Similarly, Wells-Parker, Bangert-Drowns, Mcmillen and Williams (1995) investigated the effect of quality for recidivism, but not for crashes, probably due to the small number of studies. It is therefore

still not really known how large the effects of driver improvement schemes may be, as the uneven quality of the studies has probably obscured the results too much to allow the true effect to be reliably estimated.

4.2 Limitations

The simulation used to calculate the size of the error in road safety evaluation studies in the present study may of course be criticised because the crashes which were removed were all from the culpable category. This was done under the assumption that such crashes are the only ones which can possibly be influenced by a change in driver behaviour. If this assumption is not accepted, then the calculations presented here become meaningless. However, if the assumption is not accepted, then those who take this standpoint will have to explain how safety education could possibly influence incidents where the trained driver is at a standstill (and has been for a while), and someone else crashes into the vehicle, and similar accidents. Also, several other features of culpable and non-culpable crashes, as described in af Wåhlberg and Dorn (2007), and af Wåhlberg (2008; 2009) would need some sort of different explanation.

As stated, effects of interventions can be calculated in three different ways. Apart from between- and within-subjects comparisons it is also possible to calculate an effect between before and after periods, for both intervention and control, and detract the latter from the former. In this situation, it becomes uncertain exactly what will happen concerning when using culpable accidents, as the sizes of under-estimation of effects might be different.

The current semi-simulation of the effect of using culpable crashes only instead of all crashes in evaluations is not directly transferable to the evaluations discussed, due to differences in the populations used, and some caution is therefore needed in the interpretation of the results. Professional drivers tend to be both more experienced (in years and hours of driving) and more exposed to risk than private car drivers. They will therefore probably have a higher percentage of non-culpable crashes, which make the effect investigated here larger than for private drivers (although the meta-analytic results indicated the opposite). However, a bus is also much more difficult to handle, especially in cramped and at times congested circumstances, as for the drivers used here, which will probably to some degree counter the experience and exposure effects.

4.3 Conclusions

All in all, it is not really known what the percentage of culpable crashes is in samples such as those used in driver improvement evaluations, or exactly how this influences the results. In studies which have tried to categorize according to culpability, these percentages have differed rather strongly, although most have reported much lower proportions of culpable incidents than in the present data (af Wåhlberg, 2009; af Wåhlberg & Dorn, 2007). This should indicate that the culpability effect would be larger in such samples. However, this is probably not so, because the differences are probably due to differences in how culpability has been defined (af Wåhlberg, 2009).

It has here been stated that all driver improvement evaluations using crashes apparently have under-estimated the effect. The simulations presented indicate that the difference might be somewhere between fifteen and thirty percent, depending upon what kind of effect size statistic is used. However, although the present simulation should be seen as an example of what happens when culpable crashes are used as outcome instead of all crashes, it is not certain that the effect sizes reported are very similar to those for car drivers (the most common object of driver improvement courses). More research on culpability for crashes for car drivers is therefore needed.

The importance of culpability in traffic safety research has largely been over-looked, and has never been an issue which has been discussed or researched even by the few researchers who have actually used the concept (af Wåhlberg, 2009). Although some previous attempts to study the impact of culpability on effect estimates for individual differences have been published (af Wåhlberg, 2008; 2009; af Wåhlberg & Dorn, 2007), the present paper would seem to be the first to apply this kind of logic to interventions, and calculating an estimate of how large the effect is. Hopefully, this state of the art will soon change.

Acknowledgement
The data used in paper was kindly provided by the bus company Gamla Uppsalabuss in Uppsala, Sweden. It is part of a personal database which has been used for a number of papers (e.g. af Wåhlberg, 2002; 2004; 2006; 2007; 2008).

5. References

Barraclough, P., af Wåhlberg, A. E., Freeman, J., Watson, B., & Watson, A. (2016). Predicting crashes using traffic offences. A meta-analysis that examines potential bias between self-report and archival data. *PLoS ONE.* http://dx.doi.org/10.1371/journal.pone.0153390

Christie, R. (2001). *The Effectiveness of Driver Training as a Road Safety Measure: A Review of the Literature.* Report 01/03. Melbourne: RACV.

Dorn, L., & af Wåhlberg, A. E. (submitted). A replication study of coding for culpability in traffic accidents.

Harrington, D. M. (1972). The young driver follow-up study: An evaluation of the role of human factors in the first four years of driving. *Accident Analysis and Prevention, 4,* 191-240.

Helander, C. J. (1984). Intervention strategies for accident-involved drivers: An experimental evaluation of current California policy and alternatives. *Journal of Safety Research, 15,* 23-40.

Janke, M. K. (1994). Mature driver improvement program in California. *Transportation Research Record, 1438,* 77-83.

Kaestner, N. F. (1968). Research in driver improvement: The state of the art. *Traffic Quarterly, 23,* 497-520.

Kaestner, N. (1980). *Driver Improvement Schools-An Evaluation of Two Programs.* HS-032 467. Oregon Motor Vehicles Division. Abstract. https://trid.trb.org/view.aspx?id=175942

Ker, K., Roberts, I., Collier, T., Beyer, F., Bunn, F., & Frost, C. (2005). Post-license driver education for the prevention of road traffic crashes: A systematic review of randomized controlled trials. *Accident Analysis and Prevention, 37,* 305-313.

Klein, D. (1966). A reappraisal of the violation and accident data on teen-aged drivers. *Traffic Quarterly, 20,* 502-510.

Lonero, L. P. (2008). Trends in driver education and training. *American Journal of Preventive Medicine, 35,* S316-S323.

Lonero, L., & Mayhew, D. (2010). *Large-Scale Evaluation of Driver Education. Review of the Literature on Driver Education Evaluation—2010 Update.* AAA Foundation for Traffic Safety. Washington, DC.

Lund, A. K., & Williams, A. F. (1985). A review of the literature evaluating the defensive driving course. *Accident Analysis and Prevention, 17,* 449-460.

Masten, S. V., & Peck, R. C. (2004). Problem driver remediation: A meta-analysis of the driver improvement literature. *Journal of Safety Research, 35,* 403-425.

Mayhew, D. R., & Simpson, H. M. (2002). The safety value of driver education and training. *Injury Prevention, 8(Suppl II),* ii3-8.

Mayhew, D. R., Simpson, H. M., Williams, A. F., & Ferguson, S. A. (1998). Effectiveness and role of driver education and training in a graduated licensing system. *Journal of Public Health Policy, 19,* 51-67.

McGuire, F. L., & Kersch, R. C. (1969). A study of history, philosophy, research methodology, and effectiveness in the field of driver education. *University of California Publications in Education, 19.* Berkeley: University of California Press.

Musicant, O., Lotan, T., & Toledo, T. (2007). Safety correlation and implications of an in-vehicle data recorder on driver behavior. In: Preprints of the 86th Transportation Research Board Annual Meeting, Washington, DC.

Peck, R. C. (2011). Do driver training programs reduce crashes and traffic violations? A critical examination of the literature. *IATSS Research, 34,* 63-71.

Roberts, I., Kwan, I., & Cochrane Injuries Group Driver Education Reviewers. (2008). School-based driver education for the prevention of traffic crashes (Review). *The Cochrane Library 2008, (1).*

Robertson, M. D., & Drummer, O. H. (1994). Responsibility analysis: a methodology to study the effects of drugs in driving. *Accident Analysis and Prevention, 26,* 243-247.

Rongen, A., Robroek, S. J., van Lenthe, F. J., & Burdorf, A. (2013). Workplace health promotion: A meta-analysis of effectiveness. *American Journal of Preventive Medicine, 44,* 406-415.

Strathman, J. G., Kimpel, T. J., & Leistner, P. R. (2007). *Evaluation of the Oregon Driver Improvement Program.* Urban Studies and Planning Faculty Publications and Presentations. Paper 138.

Struckman-Johnson, D. L., Lund, A. K., Williams, A. F., & Osborne, D. W. (1989). Comparative effects of driver improvement programs on crashes and violations. *Accident Analysis and Prevention, 21,* 203-215.

Vernick, J. S., Li, G., Ogaitis, S., MacKenzie, E. J. Baker, S. P., & Gielen, A. C. (1999). Effects of high school driver education on motor vehicle crashes, violations, and licensure. *American Journal of Preventive Medicine, 16,* 40-46.

Wells-Parker, E., Bangert-Drowns, R., Mcmillen, R., & Williams, M. (1995). Final results from a meta-analysis of remedial interventions with drink/drive offenders. *Addiction, 90,* 907-926.

af Wåhlberg, A. E. (2002). Characteristics of low speed accidents with buses in public transport. *Accident Analysis and Prevention, 34,* 637-647.

af Wåhlberg, A. E. (2003). Some methodological deficiencies in studies on traffic accident predictors. *Accident Analysis and Prevention, 35,* 473-486.

af Wåhlberg, A. E. (2004). Characteristics of low speed accidents with buses in public transport. Part II. *Accident Analysis and Prevention, 36,* 63-71.

af Wåhlberg, A. E. (2006). Speed choice versus celeration behavior as traffic accident predictor. *Journal of Safety Research, 37,* 43-51.

af Wåhlberg, A. E. (2007). Long term effects of training in economical driving; fuel consumption, accidents, driver acceleration behavior and technical feedback. *International Journal of Industrial Ergonomics, 37,* 333-343.

af Wåhlberg, A. E. (2008). The relation of non-culpable traffic incidents to bus drivers' celeration behavior. *Journal of Safety Research, 39,* 41-46.

af Wåhlberg, A. E. (2009). *Driver Behaviour and Accident Research Methodology; Unresolved Problems.* Farnham: Ashgate.

af Wåhlberg, A. E., Barraclough, P., & Freeman, J. (2016). Personality versus traffic accidents; meta-analysis of real and method effects. *Transportation Research Part F: Traffic Psychology and Behaviour, 44,* 90-104.
http://dx.doi.org/10.1016/j.trf.2016.10.009

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.

Wundersitz, L. N., & Hutchinson, T. P. (2006). *South Australia's Driver Intervention Programme: Participant Characteristics, Best Practice Discussion and Literature Review.* CASR021, Centre for Automotive Safety Research, Adelaide.

Zhu, M., Cummings, P., Chu, H., Coben, J. H., & Li, G. (2013). Graduated driver licensing and motor vehicle crashes involving teenage drivers: an exploratory age-stratified meta-analysis. *Injury Prevention, 19,* 49-57.
doi:10.1136/injuryprev-2012-040474



https://dspace.org/

The effect of driver improvement interventions on crash involvement; has it been under-estimated?

af Wåhlberg, Anders E.

2018-03-05
Elsevier
https://doi.org/10.1016/j.trf.2018.02.027
af Wahlberg AE. (2018) The effect of driver improvement interventions on crash involvement; has it been under-estimated? Transportation Research Part F: Traffic Psychology and Behaviour, Volume 54, April 2018, pp. 349-356

Downloaded from a DSpace repository. https://dspace.org/