

How reliable are self-report measures of mileage, violations and crashes?

*A. E. af Wåhlberg (1)**
and
L. Dorn (2)

1. Department of Psychology
Uppsala University
P. O. Box 1225
751 42 Uppsala
Sweden

2. School of Aerospace, Transport and Manufacturing
Cranfield University,
Cranfield,
Bedfordshire,
MK43 0AL
UK

* corresponding author

Safety Science (In Press June 2015)

Abstract

The use of self-reported driver mileage, violations and crashes is very popular in traffic safety research, but their validity has been questioned. One way of testing validity is with an analysis of test-retest reliability. Three mechanisms might influence reliability in self report; actual changes in the variable over time, stable systematic reporting bias, and random error. Four samples of drivers who had responded twice to an online questionnaire asking them to report their mileage, violations and crashes were used and correlations between self reports for this data were calculated. The results for crashes were compared to expected correlations, calculated from the error introduced by the non-overlapping periods and the variable means. Reliability was fairly low, and controlling for mileage in the violations and crashes calculations did not strengthen the associations. The correlation between self reports of crashes in different time periods was found to be much larger than expected in one case, indicating a report bias, while the other correlation agreed with the predicted value. The correlations for overlapping time periods were much smaller than expected. These results indicate that drivers' self reports about their mileage, violations and crashes are very unreliable, but also that several different mechanisms are operating. It is uncertain exactly under what circumstances different types of self report bias is operating. Traffic

safety researchers should treat the use of self-reported mileage, violations and crashes with extreme caution and preferably investigate these variables with the use of objective data.

Key words: common method variance, self-report, reliability, penalty points, mileage

1. Introduction

For many decades, it has been common practice in research on individual differences in traffic safety to use self-reported data for crashes, violations and mileage, assuming that drivers can and will report this information in a reliable way. Given the difficulties of acquiring official data, and the problematic properties of these sources of information (for a review see Hauer & Hakkert, 1988), many researchers have concluded that official data are inferior to self-reports (e.g. McGuire, 1973; Panek et al., 1978; Malta et al., 2005), or ignored the question and used self-reports regardless of this issue. Consequently, research on individual differences in traffic safety has mainly been built on a foundation of self-reported data.

Many traffic safety researchers argue that if an effect is found in self-report data, despite the doubtful validity of self reporting, the real effect is probably even stronger (see the review in af Wåhlberg, 2009). It is assumed that any reporting biases will only affect the independent variables, and therefore cause random error at worst, which will only serve to detract from the real effects. However, the same biases can also influence the outcome variables (mainly crashes and violations). Recently, such systematic biases in self-reports of crashes have been reported (af Wåhlberg et al., 2010; see the review in af Wåhlberg, 2009) as well as common method variance (artefactual associations in the data) (af Wåhlberg, 2009; 2010a; af Wåhlberg et al., 2011).

As with common method variance, the most obvious reason why traffic safety researchers have neglected to consider or disregarded the possibility that self-reported crashes and violations could be systematically biased may be because their properties have not been tested. Although many studies have reported extremely low agreement between state records and self-reports for these variables, this has usually been blamed upon the low validity of the records (see the review in af Wåhlberg, 2009). It would therefore be prudent to examine the psychometric property of reliability for self reported crashes and violations, and their closely related counterpart mileage, as assessing reliability can tell us something about the size and nature of the bias involved. However, as will be discussed, the assessment of reliability for these variables is probably more complicated than previously anticipated.

Turning back to traffic safety research, there are two variables that are commonly used as outcome parameters; crashes and violations (also called offences, citations, endorsements, penalties etc.). These are often used in conjunction with mileage, with the latter as a control for exposure. Given that the time periods for canvassing self-reports of mileage, violations and crashes are most often over a period of years (i.e. asking how many crashes the participant has had in the last 3 years is common in the traffic safety literature; af Wåhlberg, 2003), it seems fair to suggest that several hundred researchers have assumed that drivers report these data consistently over time periods of at least a few months.

However, tests of consistency of reporting are difficult to find for mileage, violations and crashes (af Wåhlberg, 2009). Arthur (1991) and Arthur and Graziano (1996) reported very high correlation coefficients between repeated self-reports of crashes, but as there were only two or three days between measurements (i.e. the report periods were strongly overlapping), anything but a perfect association is somewhat suspect.

Turning to correlations of self-reported crashes for different time periods, an anomaly exists in the literature. In French et al. (1993), a correlation of .305 between crashes in a period of three years versus one year was reported. As will be described below, this value would seem to be strongly inflated. Furthermore, in West et al., (1992), the same data were apparently used, but the correlations reported were much lower (see the discussion in af Wåhlberg, 2009). No other studies of the reliability of self-reported accidents have been found.

Concerning self-reported violations, no tests of reliability over time have been located.

However, some results have been published which may indicate reporting bias. Apart from generally low correlations between state sources and self-reported violations, it has also been found that drivers tend to report many more violations than those found in official records.

Whether this is due to a reporting bias, record cleaning practices and/or self-reporting of violations in other States (many of the studies were from the US) is not known (af Wåhlberg, 2009).

For mileage, the most relevant study was authored by Alonso et al. (2006), who found that self-reports correlated .64 over a period of ten months (other tests indirectly indicate low reliability due to computational difficulties, where people give different estimates due to the time period used, see the review in af Wåhlberg, 2009). Here, the method of measurement differed between occasions, with questionnaire and interview data being collected. More importantly, however, was that the instruction was to report mileage for a specific period of time (i.e. total overlap of report periods), instead of the last time period before the question was asked. For short time periods, as those studied by Arthur et al., this difference is probably of no importance, but as the period lengthens, the reports will differ due to changes occurring naturally over time.

Considering mileage alongside violations and crashes, it can be noted that exposure (conceptualized in the variable mileage) is most often seen as having a positive and causal

association with these variables, but otherwise unrelated to what drivers carry with them into the driving situation (i.e. individual differences). This is never explicitly stated, but can be concluded from how the mileage variable is treated within traffic safety research, i.e. it is controlled for as a confounder (see further the discussion in af Wåhlberg, 2009). However, contrary to this concept and use of the mileage variable in practice, controlling for it does not yield larger effect sizes (af Wåhlberg, 2009), i.e. it is usually not a confounder in the sense of hiding an effect. Instead, it may carry some of the effect when crashes and violations are predicted from other variables (af Wåhlberg, Barraclough & Freeman, submitted). No study has been located which has controlled for the influence of mileage for between time period correlations for crashes and violations.

Within psychometrics, reliability (i.e. similarity in results between measurements) of the variables measured is considered to be of importance (Anastasia, 1988), due to the basic logic that an unstable variable cannot be predicted (there are also other reasons that are not of any concern for the present paper). Therefore, if self-reported crashes and other traffic safety indicators are found to have low reliability, this reflects on the evaluation of results where significant associations have been found for these variables in relation to predictors. If reliability is low, results may be an artefact, because an unreliable measure cannot predict anything, or be predicted, unless there is some sort of bias that creates an association.

The social sciences have attempted to explain why people under- or over-report which may have a bearing on the accuracy of self reported traffic safety indicators. The research points to either recall issues due to the self report time period between the event and subsequent recall (Anderson & Schooler, 1991) and/or distorted memories. It could be that response styles which are relatively stable over time (He & van de Vijver, 2014) may reflect self-presentational needs when communicating information about oneself to others (Paulus, 1991). Unreliability in self reported traffic safety indicators may also be a function of distorted memories. Crashes and penalty points are negative events that may be more prone to memory distortion (Porter et al., 2010) but disentangling faking behaviour from genuine response errors is difficult (Shoss & Strube, 2011). If specific information is not readily accessible in memory, respondents may draw from their belief about their standing and in the general domain (Willard & Gramzow, 2008) and drivers tend to be optimistically biased about their driving skills and abilities (McKenna, 1993).

For young drivers in particular, it appears that their self esteem is inextricably linked with self perceptions of themselves as drivers (Falk & Montgomery, 2007) including a particular bias in their beliefs about their driving skills and chances of being involved in a crash relative to older drivers (e.g. McKenna, Stanier & Lewis, 1991). These kind of cognitive biases may distort the memories amongst young people in a systematic way and impact on the reliability of self reported crashes and driving offences.

The standard method for reliability measurements is to calculate the correlation between the results from different data collection occasions (Anastasi, 1988). However, usually no effort is made to investigate the possible determinants of the changes between measurements. This is probably due to an assumption that the phenomena to be studied are fairly stable over time, and that the error involved is random.

However, psychometric constructs are usually continuous and omnipresent, while crashes and violations are discrete events in time. Therefore, when calculating and interpreting reliability for these variables, there are different factors that must be taken into account. This difference does not seem to have been noted by those few researchers who have reported test-retest correlations for these variables.

When discrete event type data is used for reliability calculations, and is objective and valid, then reliability indicates the stability of the phenomenon in itself over time. For example, actual, recorded, traffic collision involvement has a certain between-time periods correlation, which increases strongly with the extension of the periods (af Wåhlberg & Dorn, 2009; see also the meta-analysis in af Wåhlberg, 2009). A similar effect can be expected for violations, although this cannot be meta-analytically tested, due to lack of published results.

However, if crash involvement is self reported, there may be two different types of report bias with opposing effects. First, a stable systematic difference in reporting tendency may exist. For example, drivers with many crashes on record tend to under-report their involvement (see the review in af Wåhlberg, 2009), but there are also drivers who over-report (af Wåhlberg, 2002) although this phenomenon is in need of replication. Furthermore, report bias has been identified by the use of a lie scale and comparisons of effects versus self-reported and recorded crash data (af Wåhlberg et al., 2010; af Wåhlberg, 2011a). Therefore, it can be suspected that self-reported data may inflate reliability, due to individual differences in reporting bias.

Second, random error may operate, due to memory difficulties. It has been shown that about 25-30 percent of collisions are forgotten per year (Maycock et al., 1991; Maycock & Lester, 1995; af Wåhlberg, 2012). If the memory effect is unsystematic i.e. not due to poor recall for some drivers, but the result of environmental circumstances, this should have a detrimental effect on the between-measurements reliability of self-reported crashes. The memory effect as

studied so far only indicates forgetting over time. However, another mechanism is also possible. Memories for events might fluctuate somewhat over time. This would imply that at one time an incident is remembered, while at another time it is not. Therefore, drivers may report crashes at Time 2 that they failed to recall at Time 1 (as found by Alonso et al., 2006). Thus, random reporting errors are added to the actual unreliability of a driver's accident record for different periods. Given the above, for the present study, it could not be predicted whether the correlations found between self-reports would be larger or smaller than expected, because there are two competing forces which could influence the results.

When calculating reliability for self-reported crash and violation data over time, we can distinguish between two different reliabilities; first, test-retest reliability of reports for the same time period, and second, reliability between different time periods. Using the first method, the natural changes between measurements are excluded, while in the second the actual stability of the variable is the starting point, upon which the two effects discussed above will have their influence.

Turning to the matter of how to determine what kind of mechanism is operating in self-reported crash and violation data, it can be noted that no criterion has ever been stated for the different kinds of reliability coefficient that can be expected in a calculation. For crashes, this lack of knowledge can be addressed in the present paper.

A new method for testing whether the inter-correlation between crashes in different time periods is different from what can be expected is to calculate the expected correlation from the mean of crashes in the sample and compare it with the actual value found. This method is based upon the finding that correlations between crashes in different time periods is strongly determined by the mean, and that the increase is linear within the time periods such as those commonly used within traffic safety research (af Wählberg, 2003; 2009). Therefore, an

expected correlation can be calculated using the regression equation for the correlation between time period (squared) correlation and the mean of crashes reported.

When, on the other hand, reports are for the same time period, the correlation should be perfect ($r=1.0$). However, when self reported crashes/violations are for a certain time period *before* the measurement occasion, the reporting periods will differ according to the time difference between the questionnaire waves, and a natural difference in reports should be entered. This expected difference can also be calculated, because it can be estimated how highly self reports for overlapping time periods should correlate if there is no reporting error, but only natural differences, due to the time difference between the information being requested. This can be calculated from the number of events in the samples and the time differences between reports as most participants have zero or one crash. The assumption made is that crashes are evenly distributed over the reporting periods. Therefore, the number of crashes happening in the periods that are not overlapping can be calculated, and compared to the differences in reporting. It should be noted that this method yields an over-estimation of the natural change in number of crashes, because it does not take into account the possibility that a driver with a crash in the first non-overlapping period might have a crash in the second non-overlapping period.

Thereafter, the (maximum) difference in crash number between measurements, due to the non-overlapping time periods, can be entered into a new variable. In other words, the first measurement is duplicated, and the crashes that can be expected to have taken place in the first non-overlapping time period are removed, and the crashes in the second non-overlapping period added, both randomly. Between the first, actual measurement, and the 'simulated' one, an expected correlation can be computed, to which the actual result between measurements can be compared.

As noted, there are two competing self report effects which may be operating and possibly influencing the results of the reliability calculations for crashes and violations. Given the two different calculation methods for reliability described (different and overlapping time periods), what results can be expected? With little previous research and no theory from which to make specific predictions, the only consideration is the nature of the difference between the two tasks being undertaken.

In the overlapping time period situation, the participant is required to report the same information twice. This means that report bias will, in essence, be a constant, while the memory effect will have its full effect, although the time periods between these in the present study were rather short. In the different time period task, the report bias will have its full effect.

In summary, the present study set out to investigate how self reported mileage, violations and crashes are associated with these same variables reported at another time point. The crash correlations were compared with expected values, which were derived from what was known about how this variable should behave in the presence of no reporting error.

2. Method

2.1. Samples and variables

The data used in the present study were gathered in three different evaluation studies concerning the effects of e-learning based driver education on the behaviour of traffic offenders in the UK (af Wählberg, 2010b; 2011b; 2013). Participants from an educational scheme for young offending drivers (Young Driver Scheme, YDS) who responded to the first and third waves of a questionnaire were included in the present study, as well as a random control sample from an e-mail campaign in the same project. These two samples responded to

questionnaire items relating to crashes and violations for different time periods while mileage period was not specified (see the Appendix for the items).

A sample of traffic offenders caught not wearing their seatbelt taking part in an online educational Seatbelt Scheme (SS) was also used for the present study. Finally, a sample of drivers undertaking a course on the dangers of going through traffic lights on red referred to as the Red Light Scheme (RLS) was used. These drivers responded to two waves of the same questionnaire, i.e. the questions relating to mileage, violations and crash items were exactly the same.

Drivers committing relevant violations in the areas implementing these e-learning based driver offender schemes could choose to pay a fine and receive penalty points on their driving license (except the seatbelt offenders), or pay for and complete the e-learning course on traffic safety related to their offence. More than ninety percent of the offenders agreed to take part in the YDS and SS, while for RLS this was less than 50%. The remaining drivers in RLS elected to do a classroom-based course or pay the fine and accept the penalty points on their license. The re-response rates of the questionnaire last waves were fairly low for all samples (4.3 to 16.0 percent), but this was not considered relevant for the present type of analysis. Also, the differences between those who responded to only the first wave (virtually all course takers) and the present responders have been shown to be surprisingly small (af Wåhlberg & Poom, submitted).

2.2 Procedure

All first-wave questionnaires were completed online in conjunction with the first e-learning module of the course. For the final wave (second or third dependent on the nature of the evaluation study design), an e-mail was sent requesting that the driver go online and respond again to the same questionnaire. The Control sample respondents were sent an e-mail asking

them to participate in an online survey on driving behaviour, and had the opportunity of winning a GPS navigator as an incentive to participate. Those who responded to the first wave were sent an e-mail six months later, asking them to respond again.

2.3 Analysis

For the samples where crashes were reported for different time periods (YDS and Control), it could be estimated how strong the correlation would actually be from the sample mean (af Wåhlberg, 2009). In essence, the inter-correlation increases with an increased mean number of crashes, and about eighty percent of the variance in between time periods correlation size can be explained in this way. Thus, using the regression equation from the meta-analysis reported in af Wåhlberg (2009), the amount of explained variance (squared correlation) of crashes between time periods, can be predicted with very good precision, using mean number of crashes in the sample as a predictor¹.

The predicted correlation thus represents the most probable effect size, given the sample mean (both waves added). If the actual correlation in self-reported data is higher than the predicted value, this would indicate that reporting bias is the stronger factor. A lower value tells us that random error is stronger, and the difference is an estimate of the smallest possible effect of the variable in question.

Furthermore, for the overlapping time periods samples (SS and RLS) it was possible to calculate an expected lowest correlation which took the natural change in the number of crashes over time into account, under the assumption that crashes were evenly distributed over time. If so, it could be calculated how many crashes took place in the period that was not covered by the other report, and the difference subtracted.

¹ The formula used in the present study was the one calculated for car drivers: $r^2=0.002+0.019*\text{Mean of crashes in sample}$, as given in af Wåhlberg (2009), Table 3.4. Formulas for bus and tram drivers were slightly different.

For the violation data, no expected correlation for different time periods could be calculated, as no meta-analysis for such data was available. For the overlapping period, this could in principle have been undertaken, but there is a problem in determining what kind of numbers might be expected, as violations were reported in penalty point category form, and there was not an exact agreement between these numbers and number of violations.

Although no comparative values could be estimated for mileage the correlation should be fairly high, given the wording of the question and the short time periods between measurements.

3. Results

Descriptive results for the samples can be seen in Tables 1 and 2. No extraordinary values were apparent in these calculations, possibly apart from the over representation of men in the seatbelt scheme sample.

Tables 1-2 about here

Thereafter, mileage was correlated with crashes and violations (Table 3). As is commonly found, these associations were rather weak.

Table 3 about here

As can be seen in Table 4, the expected differences in number of crashes were very different from the actual numbers. Taking the SS sample as an example, the time period in the first wave not covered by the second report was three months. This means that eleven crashes

would have happened in the period not covered by the second report. As some of these may have been replaced by new events in the last part of the second report period, this is an overestimate of the expected difference. However, as can be seen in Table 4, as many as forty-five crashes reported in the first wave were not reported in the second wave. Similarly, for the second wave, fifty-five collisions were reported that had not been reported before, when twelve crashes would have been expected. The large discrepancies between the self reports can therefore not be explained to any high degree by the difference in time periods. In the next step of the analysis, the expected differences were subtracted/added to a copy of the first self report variable in each sample, thus creating a 'simulated' second response. The first and simulated variables were correlated, yielding an expected lowest possible correlation, which is presented in Table 5.

Table 4 about here

Turning to the actual reliability computations (see Table 5), it was found that mileage and penalty points correlations between waves were fairly high according to social science standards, but still did not explain more than about half of the variance.

The crash data yielded a more complex picture than the other variables. In the different time period situation, correlations were very low, while in the overlapping conditions they were medium-sized. Comparing the predicted values to the actual correlations found, it can be seen that the Control sample results matched the predicted value well, while the YDS explained more than three times as much variance. The SS and RLS samples had substantially lower correlations than expected. Controlling for mileage had a negative effect in each instance.

Table 5 about here

The fairly strong correlation for crashes per mile in the RLS sample was mainly caused by two outliers. Without these outliers, the correlation reduced to .387. For the Control group crashes per mile correlation, one outlier was excluded which otherwise created a strong positive correlation. For the penalty points per mile association, there were seven outliers, but if these were removed, the correlation was still very low. In general, calculations of correlations with crashes per mile are often strongly influenced by extreme values, and the present problems were therefore to be expected.

4. Discussion

The present results would seem to imply that self-reports of crashes, violations and mileage are very unreliable, given the low correlations between measurements, even though the time periods between measurements were fairly short, as compared with the length of time such predictors are usually reported for (af Wählberg, 2003). This effect was especially evident when the drivers should have been mainly reporting the same events on both occasions. The question is why this is so.

The finding that the YDS inter-correlation between crashes in different time periods was much stronger than could be expected from its mean is an indication that some sort of reporting tendency was operating in this sample, while for the Control group, the effect corresponded with the prediction. This result is in agreement with previous analyses, in which the YDS sample was strongly influenced by social desirability, while the effect was weaker for the Controls (af Wählberg, 2010a; 2010b).

Does the correspondence between the expected and actual values for the Control sample mean that no response bias was operating here, and no random error? Although this is of course

possible, it is equally possible that both were at play, counteracting each other. However, as the present study could not tease these possible effects apart, it is impossible to know.

Given the limits of the design, it is therefore difficult to draw any firm conclusions about how strong systematic response bias and random error effects are in crash data, not to mention self reported violations and mileage. What can be concluded with certainty is that peculiar and strong effects can be found in such data, and that the problems of reliability and its determinants deserve greater scrutiny from traffic safety researchers.

Some limitations of the present results are worthy of notice. First, most of the samples were mainly collected from the traffic offender population, which probably make them somewhat different from drivers in general. However, these traffic offenders are not extreme in any way. On the contrary, the traffic offences they committed were of the types that are very common (low level speeding, red light running, driving without seatbelt; Porter & England, 2000; Porter & Berry, 2001; Streff & Wagenaar, 1989; Parada et al., 2001), and are offences that are committed by most of the driving population at some time. The samples used in the present study may therefore be seen as rather normal drivers. However, as some of the participants were rather young, it is possible that they were more vulnerable to providing unreliable self reports due to specific cognitive biases about their driving skills and crash risk (McKenna, Stanier and Lewis, 1991) or memory distortion due to the time period between the traffic event and subsequent recall.

More importantly, the participants experienced a situation that was rather different from that of most questionnaire responders, as they were under the scrutiny of the police, with the questionnaire to evaluate their education being closely associated with this situation. This has been shown to have strongly affected the drivers' responses, with very low levels of socially presented behaviors before the course, while after the course this was not so (af Wählberg,

2010b; 2013). Whether this difference between waves could influence the present results is not known.

In previous tests of the reliability of self reported crashes, violations and mileage over time, no effort seems to have been made to disentangle the various possible effects involved; reporting bias, random error (mostly memory problems) and natural changes in the values reported (e.g. Arthur, 1991). In the present study, two new methods were devised for the crash variable, based upon the special properties of the variable and previous research into the reliability of recorded accident data. The interpretation of the differences between actual and expected correlations thus rests on the acceptance of these new methods for calculating the expected correlations. If these methods are not found to be valid, all that can be concluded from the present results is that reliability of self-reported driver events data is very low.

Returning to the original research objective, it can be asked what the present results mean for common method variance in traffic safety research? As stated in the introduction, reliability restricts the possibility of prediction. Therefore, how can an independent variable with low reliability predict a dependent variable with even lower reliability? The actual effect would need to be exceedingly strong to come through such error-laden data.

Such an interpretation would be welcomed by many traffic safety researchers, as it is exactly what they have stated when considering the problem of the low or in most cases unknown, reliability of their predictors (see the review in af Wählberg, 2009). However, such a statement would seem to run into another difficulty. There are many different researchers that would like to claim strong effects for their crash predictors (see for example the meta-analyses by de Winter & Dodou, 2010; Clarke & Robertson, 2005). If you add these together, you will probably end up with more than one hundred percent explained variance, meaning that each factor cannot be claimed as 'the' predictor, because they are intercorrelated and/or

the effect sizes are inflated, due to common method variance. There are simply too many strong effects claimed by too many.

Acknowledgements

The data used in this paper were gathered in projects involving the police forces of Thames Valley and Greater Manchester, and the companies DriveTech and a2om. Chris Johnson (a2om) set up the questionnaires on line.

References

Alonso, A., Laguna, S., & Seguí-Gomez, M. (2006). A comparison of information on motor vehicle crashes as reported by written or telephone interviews. *Inj Prev, 12*, 117-120.

Anastasi, A. (1988). *Psychological Testing*. New York: Macmillan.

Anderson, J. R. & Schooler, L. J. (1991). Reflections of the environment on memory. *Psych Sci, 2*, 396-408.

Arthur, W. Jr. (1991). *Individual differences in the Prediction and Training of Complex Perceptual-Motor Skill Tasks: The Development and Validation of The Computer-Administered Test of Visual Selective Attention*. Technical report. Texas A&M University.

Arthur, W., Jr., & Graziano, W. G. (1996). The Five-factor model, conscientiousness, and driving accident involvement. *J Pers, 64*, 593-618.

Clarke, S., & Robertson, I. T. (2005). A meta-analytic review of the Big Five personality factors and accident involvement in occupational and non-occupational settings. *J Occup Organ Psych*, 78, 355–376.

Falk, B., & Montgomery, H. (2007). Developing traffic safety interventions from conceptions of risks and accidents. *Transp Res Part F*, 10, 414-427.

French, D. J., West, R. J., Elander, J. & Wilding, J. M. (1993). Decision-making style, driving style, and self-reported involvement in road traffic accidents. *Ergon*, 36, 627-644.

Hauer, E., & Hakkert, A. S. (1988). Extent and some implications of incomplete accident reporting. *Transp Res Rec*, 1185, 1-10.

He, J & van de Vijver, F. J. R (in press). Self presentational styles in self reports: Linking the general factors of response styles, personality traits and values in a longitudinal study. *Pers and Indiv Diff*.

Malta, L. S., Blanchard, E. B., & Freidenberg, B. M. (2005). Psychiatric and behavioral problems in aggressive drivers. *Behav Res Ther*, 43, 1467-1484.

Maycock, G. & Lester, J. (1995). Accident liability of car drivers: follow-up study. In G. B. Grayson (Ed.) *Behavioural Research in Road Safety V*, pp. 106-120. Crowthorne: Transport Research Laboratory.

Maycock, J., Lockwood, C., & Lester, J. F. (1991). *The Accident Liability of Car Drivers*. TRRL Research Report No. 315. Crowthorne: Transport and Road Research Laboratory.

McGuire, F. L. (1973). The nature of bias in official accident and violation records. *J Appl Psych*, 57, 300-305.

McKenna, F. P., (1993). It won't happen to me: Unrealistic optimism or illusion of control? *Brit J Psy*, 84, 39-50.

McKenna, F. P., Stanier, R. A., & Lewis, C. (1991). Factors underlying illusory self assessment of driving skill in males and females. *Acc Anal Prev*, 23, 45-52.

Panek, P. E., Wagner, E. E., Barrett, G. V., & Alexander, R. A. (1978). Selected Hand test personality variables related to accidents in female drivers. *J Pers Assess*, 42, 355-357.

Parada, M. A., Cohn, L. D., Gonzalez, E., Byrd, T., & Cortes, M. (2001). The validity of self-reported seatbelt use: Hispanic and non-Hispanic drivers in El Paso. *Acc Anal Prev*, 33, 139-143.

Paulus, D. L., (1991). Measurement and control response bias. In J. Robinson, P. Shauer and L. Wrightsman (Eds). *Measures of Personality and Social Psychological Attitudes* (Vol 1, pp. 17-59). San Diego, CA: Academic Press.

Porter, B. E., & Berry, T. D. (2001). A nationwide survey of self-reported red light running: measuring prevalence, predictors, and perceived consequences. *Acc Anal Prev*, 33, 735-741.

Porter, B. E., & England, K. J. (2000). Predicting red-light running behavior: A traffic safety study in three urban settings. *J Saf Res*, 31, 1-8.

Porter, S. et al. (2010). A prospective investigation of the vulnerability of memory for positive and negative emotional scenes to the misinformation effect. *Can J of Beh Sci*, 42, 55-61.

Shoss, M, K., & Strube, M. J. (2011). How do you fake a personality test? An investigation of cognitive models of impression-managed responding. *Org Beh and Hum Decision Process*, 116, 163-171

Streff, F. M., & Wagenaar, A. C. (1989). Are there really shortcuts? Estimating seat belt use with self-report measures. *Acc Anal Prev*, 21, 509-516.

West, R., Elander, J., & French, D. (1992). *Decision Making, Personality and Driving Style as Correlates of Individual Crash Risk*. TRRL Contractor Report 309. Crowthorne: Transport and Road Research Laboratory.

Willard, G., & Gramzow, R.H., (2008). Exaggeration in memory: Systematic distortion of self-evaluative information under reduced accessibility. *J Ex Soc Psy*, 44, 246-259.

de Winter, J. C., & Dodou, D. (2010). The Driver Behaviour Questionnaire as a predictor of accidents: A meta-analysis. *J Saf Res*, 41, 463-470

af Wåhlberg, A. E. (2002). On the validity of self-reported traffic accident data. Manama: E140 Proceedings of the 2nd Safety On Road International Conference.

af Wåhlberg, A. E. (2003). Some methodological deficiencies in studies on traffic accident predictors. *Acc Anal Prev*, 35, 473-486.

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

af Wåhlberg, A. E. (2010a). Social desirability effects in driver behavior inventories. *J Saf Res*, 41, 99-106.

af Wåhlberg, A. E. (2010b). Re-education of young driving offenders; effects on self-reports of driver behavior. *J Saf Res*, 41, 331-338.

af Wåhlberg, A. E. (2011). The accident-exposure association: Self-reported versus recorded collisions. *J Saf Res*, 42, 143-146.

af Wåhlberg, A. E. (2011). Re-education of young driving offenders; effects on recorded offences and self-reported collisions. *Transp Res Part F: Traf Psych Behav*, 14, 291-299.

af Wåhlberg, A. E. (2012). Memory effects in self-reports of crashes. In L. Dorn (Ed.) *Driver Behaviour and Training, Volume V*, pp. 283-288. Fifth International Conference on Driver Behaviour and Training. Paris November 29-30, 2011. Farnham: Ashgate.

af Wåhlberg, A. E. (2013). Evaluation of an e-learning seatbelt wearing intervention. In M. Hamada (Ed.) *E-learning New Technology, Applications and Future Trends*, pp. 277-295. New York: Nova Science Publishers.

af Wåhlberg, A. E., Barraclough, P., & Freeman, J. (submitted). The Driver Behaviour Questionnaire as accident predictor; a methodological re-meta-analysis.

af Wåhlberg, A. E., & Dorn, L. (2009). Bus driver accident record; the return of accident proneness. *Theor Iss Ergon Sci*, *10*, 77-91.

af Wåhlberg, A. E., Dorn, L., & Kline, T. (2010). The effect of social desirability on self reported and recorded road traffic accidents. *Transp Res Part F*, *13*, 106-114.

af Wåhlberg, A. E., Dorn, L., & Kline, T. (2011). The Manchester Driver Behaviour Questionnaire as predictor of road traffic accidents. *Theor Iss Ergon Sci*, *12*, 66-86.

af Wåhlberg, A. E., & Poom, L. (submitted). An empirical test of non-response bias in surveys.

Appendix: The questionnaire items

The Young Driver Scheme (offenders and controls)

First wave

43. Approximately how many miles do you drive per month?*
44. How many collisions have you had in total as a driver while holding a full driver's license?
45. How many penalty points do you currently have on your license?

Last wave

43. Approximately how many miles do you drive per month?
46. How many collisions have you had the last six months as a driver?
47. How many penalty points have you received the last six months?

The Seatbelt Scheme, both waves, and the Red Light Scheme, first and third waves

25. Approximately how many miles do you usually drive per month?

Categories: 100, 200, 400, 600, 800, 1000, 1200, 1400, 1600, 1800, 2000 or more

26. How many collisions have you had, as driver of a motorized vehicle, in the last three years?

Please include any mishap where there was damage and/or injuries, regardless of responsibility.

Categories: (0, 1, 2, 3, 4+)

28. How many penalty points do you currently have on your license?

Categories: 0, 1-3, 4-6, 7-9, 10 or more

29. Your age group

Categories: 16-20, 21-25, 26-30, 31-35, 36-40, 41-45, 46-50, 51-55, 56-60, 61-65, 66-70, 71-75, 76 or older

31. How many years have you had a full driver's license?

Categories: 0-5, 6-10, 11-15, 16-20, 21-25, 26-30, 31-35, 36-40, 41-45, 46-50,

* This free format was quickly changed to the same set of categories that were used in the other questionnaires, due to responses that were not numbers.

Table 1: Descriptive data for the samples. Shown are the number of respondents, means and standard deviations for age (in years) and experience (years since licensing), and the percent of men in the sample.

Sample	N	Age	Experience	Sex
YDS	1152	22.2/2.2	3.69/2.20	52.4%
YDS Control	234	31.8/13.3	13.5/13.2	53.0%
SS	407	41.2/14.5	21.0/14.0	83.8%
RLS	271	42.6/12.6	21.5/12.4	57.0%

Table 2: Descriptive data for the samples. Shown are the number of respondents, the time difference between self-reports (time 1 and time 2), mileage per month (time 1 and time 2), and the mean and standard deviation of number of crashes and penalty points in the reporting periods (time 1 and time 2).

Sample	N	Time difference	Mileage/mo nth 1	Mileage/mo nth 2	Reporting periods for crashes	Crashes 1	Crashes 2	Penalty points 1	Penalty points 2
YDS	1173	7 months	643/595	591/452	Different (variable/6 mo)	0.259/0.971	0.189/0.754	0.578/1.571	0.208/0.865
YDS Control	234	6 months	497/397	480/390	Different (variable/6 mo)	0.816/1.338	0.030/0.171	0.278/0.910	0.064/0.435
SS	407	3 months	817/597	828/615	Overlapping 91.7% (3 years)	0.322/0.622	0.342/0.623	0.870/1.662	0.803/1.562
RLS	271	6 months	766/584	757/588	Overlapping 83.3% (3 years)	0.303/0.581	0.358/0.621	0.679/1.362	0.675/1.445

Table 3: Results for the effect of exposure on negative driving events. Shown are the number for respondents, and the correlations between mileage and the dependent variables number of crashes and penalty points, in each sample and wave.

Sample	N	Crashes/first wave	Penalty points/first wave	Crashes/last wave	Penalty points/last wave
YDS	1173	.116***	.081**	.045	.079**
YDS	234	.217**	.174*	.041	.167*
Control					
SS	407	.013	.074	.083	.034
RLS	271	-.004	.176**	.144*	.133*

* p<.05, ** p<.01, *** p<.001

Table 4: Calculations of the differences in numbers of crashes reported in two waves of questionnaires, when the time periods for reporting were overlapping. Shown are, first, the number of subjects, then the actual numbers of crashes reported. Thereafter a calculation of what the expected, highest possible, change in reporting would be, based upon the numbers in each wave and the difference in time period for reporting. Finally, the actual differences found.

Sample/variable	N	No 1 st wave	No 2 nd wave	Expected difference 1 st wave	Expected difference 2 nd wave	Actual difference 1 st wave	Actual difference 2 nd wave
SS crashes	407	131	139	11	12	45	55
RLS crashes	271	82	97	14	16	39	24

Table 5: The correlations between self-reports of the same variables in different waves. Shown (in rows) are the number of respondents, the overlap between report periods, the length of the report periods in years, the time elapsed between questionnaire waves, and the associations between the reports for mileage, points and crashes, and variants of these per mile. Also, the expected correlation between reports of crashes is shown. For the YDS and YDS Control, this expected correlation was based upon the means of the accident variables, while for the SS and RLS it was the degree of overlap between reporting periods which yielded these values.

Sample	YDS	YDS Control	SS	RLS
N	1173	234	407	271
Overlap of report periods	Different	Different	Overlapping 91.7%	Overlapping 83.3%
Time periods for reports	Variable/0.5 years	Variable/0.5 years	3/3 years	3/3 years
Time between waves	7 months	6 months	3 months	6 months
Mileage	.566***	.606***	.722***	.746***
Penalty points	.141*	.150*	.777***	.826***
Penalty points/mile	.018 (N=1145)	.000 (N=203)	.525***	.658***
Crashes	.216***	.137*	.562***	.632***
Expected r for crashes	.118	.134	.925	.834
Crashes/mile	.191*** (N=1145)	-.031 (N=202)	.342***	.626***

* p<.05, *** p<.001

**Published by Elsevier. This is the Author Accepted Manuscript issued with:
Creative Commons Attribution Non-Commercial No Derivatives License (CC:BY:NC:ND 3.0).**