Personality versus traffic accidents; meta-analysis of real and method effects

Running head: Personality versus traffic accidents

Abstract
Problem: The association between personality and traffic accident involvement has been extensively researched, but the literature is difficult to summarize, because different personality instruments and statistics have been used, and effect sizes differ strongly between studies.
Method: A meta-analysis of studies which had used measures of personality which could be converted into Big Five dimensions, and traffic accidents as the dependent variable, was undertaken.
Analysis: Outlier values were identified and removed. Also, analyses on effects of common method variance, type of instrument, dissemination bias and restriction of variance were undertaken.
Results: Outlier problems exist in these data, which prohibit any certainty in the conclusions. Each of the 5 personality dimensions were predictors of accident involvement, but the effects were small ($r<.1$), which is much weaker than in a previous meta-analysis. Effect sizes were dependent upon variance in the accident variable, and the true (population) effects could therefore be larger than the present estimates, something which could be ascertained by new studies using high-risk samples over longer time periods. Newer studies and those using Big Five instruments tended to have smaller effects. No effects of common method variance could be found.
Conclusions: Tests of personality are weak predictors of traffic accident involvement, compared to other variables, such as previous accidents. Research into whether larger effects of personality can be found with methods other than self-reports is needed.

Keywords: personality, accident, crash, common method variance

1. Introduction

1.1 Personality as predictor of traffic accident involvement

The present paper summarizes the literature on personality (in terms of the Big Five system) as a predictor of traffic accident involvement in a meta-analysis. Several methodological problems are considered, such as outliers, dissemination bias and conversion of data between different personality systems.

Personality as a phenomenon is multi-faceted, but can usually be defined as the stable behavioural tendencies of people over time, or the psychological traits which cause such behaviours. This has been conceptualized in many different ways through the years, but today it is agreed by most researchers that the most parsimonious description is by five dimensions; Openness, Agreeableness, Conscientiousness, Neuroticism and Extraversion. Most other systems map onto these dimensions, and results can therefore be converted between them.

Throughout the history of traffic safety, researchers have studied the influence of individual differences in personality on accident record (although at first the term 'accident proneness' was used; Greenwood & Woods, 1919; see also papers by Drake, 1940; Harris, 1950; Parker, 1953). Many researchers have proposed that certain personality features, in terms of recurrent behaviours, cause accidents. In terms of the Big Five model (and its facets), Clarke and Robertson (2005) summarized the theoretical basis for their traffic accident-causing properties thus; people high on Extraversion tend to be poor on vigilance and take more risks. Those high on Neuroticism have been suggested to be easily distracted, less likely to seek control of the environment and prone to react to stress. Conscientiousness features several inter-related concepts which are thought to make people safe, such as planning, self-control and decision-making, while lack of Agreeableness is associated with accidents by the mechanism of aggression in terms of emotion as well as behaviour. Finally, Openness has been suggested to be positively correlated with accidents, due to the routine character of driving, where traits such as experimentation and improvisation are not in accord with safe operation. However, most researchers who investigate the link between personality and accidents refer to previous significant associations reported, and describe the behaviours typical of a certain personality dimension (e.g. Arthur et al., 2001; Begg, Langley & Williams, 1999; Burns & Wilde, 1995; Clement & Jonah, 1984; Hartman & Rawson, 1992).

Many researchers also express a strong belief in the predictive capacity of tests of personality versus accidents (e.g. Arthur et al., 2001; Brandau, Daghofer, Hofmann & Spitzer, 2011; Hansen, 1988; Jonah, 1997). However, results, as always, have been mixed, and this belief may therefore be unfounded. For example, Shaw and Sichel (1971; Shaw, 1965) reported correlations between .4 and .7 for their personality tests and accidents for bus drivers, while Carty, Stough and Gillespie (1998) found a strong negative association (-.212) instead of the expected positive one for Extraversion, and many other such examples exist. Results are thus very heterogeneous, which make interpretation difficult. A meta-analytic approach is therefore needed, where the reasons for this apparent heterogeneity can be identified, and estimates of the true (population) effects calculated.

Two meta-analyses of personality versus accidents have already been published; Arthur, Barrett and Alexander (1991) and Clarke and Robertson (2005). However, there are several reasons for why a new analysis of the personality-traffic accident association is needed. Apart from now being outdated, the Arthur et al. study used a personality taxonomy which excluded some available studies (e.g. Quenault, 1967; Andersson, Nilsson & Henriksson, 1970; Jamison & McGlothlin, 1973). Similarly, the Clarke and Robertson study excluded many available papers, while including some which used methodologies which were different from those of the majority. Furthermore, moderator effects and dissemination bias were not investigated in these studies.

We therefore wanted to undertake a new meta-analysis which used a very different approach to the problem of meta-analysing personality as a predictor of traffic accident involvement, taking into account not only the well-known problems of dissemination bias and methodological moderator effects, but also effects which are probably peculiar to accident prediction studies. The main aim of the study was therefore to estimate the population effect while keeping known or suspected moderators constant, as will now be described.

1.2 Technical issues in meta-analysis; Heterogeneity and the population effect

This section describes some of the methodological problems associated with meta-analysing data, under the general headings of trying to estimate a population effect, and the overall problem of heterogeneous data, i.e. very different results in different studies. Also, possible remedies are suggested.

In research on psychological mechanisms, it is usually the goal to infer from sampled data what all people are like in a defined population. For example, are high levels of empathy usually associated with low levels of aggression? In a meta-analytic context, it would specifically be asked what the effect size is, i.e. how strong is the link between the two concepts? When effect sizes from different studies are combined, however, it is important that the data included is actually drawn from the same population, meaning those who share this trait/mechanism. For example, the link between empathy and aggression might have different strength in different cultures. If studies from different cultures are then combined, the ensuing effect size will be slightly misleading, showing really the mean effect for two (or more) different populations. When effect sizes from different populations are mixed, it is said that the meta-dataset is heterogeneous, i.e. the numbers differ more between themselves than could be expected by random sampling (which can be ascertained by statistical testing). Heterogeneity can also be caused by differences in methodology. For example, it can be expected that experiments and field studies will yield different effect sizes, although they are ostensibly studying the same problem, because part of the effect is actually created by the method used (e.g., a social science analogue to Heisenberg's uncertainty principle).

If heterogeneity is detected in the data, moderator[1] analysis should be applied to investigate the causes of the variance. For example, the pooled effects for experiments versus field studies can be compared, to see whether the estimated population effects differ significantly between these two conditions. If they do, it can be concluded that the methods used have had an influence on the results, a fact that needs to be considered when the true population effect size is identified.

When meta-data has been gathered, an important operation is therefore to detect whether effects differ more between themselves than could be expected from sampling error alone. However, before moderator analysis is undertaken, data should be checked for outlying values, i.e. values which differ very strongly from the majority, and could be suspected to be due to errors in the research process. If a few such values are found, these deviating numbers can be excluded (e.g. Bond & Smith, 1996; Eagly, Makhijani & Klonsky, 1992; de Winter & Dodou, 2010; Groh et al., 2014; Fournier, Hass, Naik, Lodha & Cauraugh, 2010), although most meta-analysts in social science do not proceed beyond concluding that there is heterogeneity between effects (while in 'hard' sciences, such as physics, they do; Hedges, 1987).

1.3 Meta-analysis of personality as accident predictor

[1] In meta-analytic jargon, a moderator is a variable which systematically influences the effect sizes in a set of studies, for example differences in methodology in research on the same problem.

In this section, the methodological problems of meta-analysis, with a special emphasis on the problems associated with accident prediction and personality, are further described. The solutions chosen for the present analysis are also described.

Variance in the accident variable has been shown to strongly affect effect sizes in accident prediction studies (af Wåhlberg, 2009; af Wåhlberg & Dorn, 2009; af Wåhlberg, Barraclough & Freeman, 2015; Barraclough, af Wåhlberg, Freeman, Watson & Watson, 2016). In essence this is a problem of differences in restriction of variance between studies. This means that if a fair comparison of effects between sources of data or other moderators is to be undertaken in accident prediction studies, variance/restriction of range should be held constant. This can be undertaken by using a correction formula suggested by Hunter & Schmidt (1990). However, this formula uses the standard deviations of the samples, a statistic which is not always reported in accident prediction studies (61% in af Wåhlberg, Barraclough & Freeman, 2015). Therefore, unless a large part of the available data is to be excluded, a different method is needed to correct for range restriction and make the results comparable between studies. In the present paper, the empirically derived association between the mean in the accident variable and the effect size for the predictor used (first calculated in af Wåhlberg, 2009) will be utilized.

The moderating effect of the variance in the accident variable also influences the statistics calculated, and the interpretation of the results. In standard meta-analysis, the goal is to calculate a mean effect over samples, and often to estimate a true population effect by correcting for measurement error (as advocated by Hunter & Schmidt, 1990). For individual differences in accident involvement, however, this is not meaningful, as any mean would only be 'true' for a specific level of variance in the accident variable[2]. Similarly, a correction for measurement error would involve the reliability of accidents, something which has not been established as a single value. Instead, a calculation of how the effect varies with the variation should be more relevant.

Yet another correction procedure which is sometimes used in meta-analysis is to adjust for the unreliability of the predictor, yielding an estimate of what the effect size would be if the predictor was perfectly measured. Given the state of the present data, it was chosen not to apply this method, as it would probably be impossible to retrieve reliability information about several of the instruments used. Instead, the main approach in this paper was to include as much data as possible, and leave correction procedures for the future, using a more restricted data set.

Errors in the research process can create very deviating values (outliers), which can unduly influence the population effect estimate (Hunter & Schmidt, 1990). However, as strong variance in effect sizes between studies due to differences in range restriction in the accident variable was expected in the present data (af Wåhlberg, 2009), standard univariate methods for outlier detection were not applicable. Instead, a new method for identifying suspect data points was applied in the present study. It uses the standard criterion of two standard deviations from the mean as a cut-off for outlying values, but applies this to bi-variate data points. This will be further described in the method section.

Dissemination bias (previously known as publication bias; Bax & Moons, 2011), such aswhen the results of a study influences its availability, is a problem in many areas of research (Ioannidis, Munafò, Fusar-Poli, Nosek & David, 2014). Therefore, many different methods for detecting such bias have been invented, mostly based upon the assumptions of larger

[2] Some statisticians recommend controlling for restricted variance in, and/or reliability of, the variable. Essentially, the method used here is equivalent to controlling for resricted variance, but without the statistical assumptions behind this method, as it uses an empirically derived formula. The test-retest reliability of the accident variable is not really known, and is mainly a function of the variance in the variable, as it explains about 80 percent of the variation between samples (af Wåhlberg, 2009).

studies having more reliable results and studies with large effects being easier to publish (Møller & Jennions, 2001). This means, for example, that if the number of subjects and the effect sizes in published studies are negatively correlated, a number of small studies with small effects have probably not been published. However, these methods are not considered fully reliable (Pham, Platt, McAuley, Klassen & Moher, 2001; Song et al., 2010; Vevea & Woods, 2005), and tend to have low power (Macaskill, Walter & Irwig, 2001; Sterne, Gavaghan & Egger, 2000) and it is therefore preferable to try to actually locate unpublished data, a method which has previously yielded a significant amount of additional data (Judge, Colbert & Ilies, 2004; Eyding et al., 2010; af Wåhlberg, Barraclough & Freeman, 2015; Barraclough, af Wåhlberg, Freeman Watson & Watson, 2016).

There are good reasons to suspect that in studies using self-reported accident data as well as self-reported predictors, effects are artificially increased (af Wåhlberg, 2009; Hessing, Elffers & Weigel, 1988; Schwartz, 1999; Podsakoff, Mackenzie, Lee & Podsakoff, 2003). The problem of single-source data, especially self-reports, is that common method variance can influence the results, and sometimes substantially change the true effect size. This distorting effect is also possible for studies into personality, leading to the prediction that studies on personality as a predictor of accident involvement will have larger effect sizes if the criterion is self-reported than if it is objectively gathered data. Such an effect has previously been found in meta-analyses by Reijntjes, Kamphuis, Prinzie and Telch (2010) for internalizing problems and peer victimization, by af Wåhlberg, Barraclough and Freeman (2015) for a driver behaviour inventory and by Barraclough, af Wåhlberg, Freeman, Watson and Watson (2016) for citations versus crashes, but see also Arthur, Barrett and Alexander (1991), and Morina, Ijntema, Meyerbröker and Emmelkamp (2015) for less clear-cut results.

In summary, the present study sets out to meta-analyse the association between personality in terms of the Big Five dimensions, measured by standard personality scales, and traffic accident involvement (main aim). All other analyses (identifying suspicious values, and testing for various method effects and biases) were included to increase the precision of the population estimate. This included effects of variance in the accident variable and common method variance, differences been inventories, as well as the more commonly known problem of dissemination bias.

2. Method

2.1 General

The preliminary work on this paper followed the standard guidelines for meta-analysis (e.g. Chung, Burns & Kim, 2006; Field & Gillett, 2010; and the discussion pieces of Noble, 2006; Aguinis, Pierce, Bosco, Dalton & Dalton, 2011; Huf et al., 2011; Orme-Johnson & Dillbeck, 2014). Thereafter, it was mainly geared towards investigation of the effect of moderators, in similarity to, for example, Bond and Smith (1996) and as described in Steel and Kammeyer-Mueller (2002). One feature that was unusual was the great number of different measures that could, in principle, be converted into the five independent variables (e.g., personality dimensions) used here, and the ensuing problems with these conversions.

2.2 Literature search

Data for the analysis was gathered by several methods. First, the meta-analysis by Clarke and Robertson (which included papers from the reviews by Hansen, 1988; Lawton & Parker, 1998, and Jonah, 1997) and other reviews (Adams, 1970; McGuire, 1976; Donovan, Marlatt & Salzberg, 1983; Lester, 1991; Beirness, 1993; Nichols, Classen, McPeek & Breiner, 2012; Signori & Bowman, 1974) were used to identify potential papers by searching their reference lists. Second, searches were made in the databases ScienceDirect, ISI, Psychinfo, PubMed and Google Scholar with the search string "personality and traffic and (accident or crash or collision)" for the last five years (2008-2012), as earlier material was considered to have been covered by the other searches. Third, the journals Accident Analysis and Prevention, Risk Analysis, Traffic Injury Prevention and Transportation Research Part F were searched with the keyword 'personality' (no time constraint). Finally, manual searches of the reference lists of all papers were undertaken.

If an e-mail address could be located, authors of papers published from 2002 and onwards that had gathered the necessary data but where some more information was needed, were e-mailed and further results requested. Apart from non-functional e-mail addresses, twenty-seven researchers were thus contacted, and eleven responded (see Acknowledgements). Of these, six could provide the data needed.

2.3 Inclusion and exclusion of studies

Included in the analysis were all studies that had tested some kind of self-report personality measure, as a predictor of the number of traffic accidents (actual crashes, not near misses). The inventory used should measure the Big Five, or permit conversion into Big Five dimensions (see section 2.5). Also, they should have reported a sufficient level of statistical detail for effects to be converted to correlations.

Studies that limited their personality measures to the driving situation only (e.g. driving aggression) were excluded. Also, the personality measures should be self-reported by the subjects in written form, thus excluding interviews and next-of-kin reports. All studies included were in English, although this was not really a criterion. All searches were made using English keywords, but none of the search methods used yielded any paper in a different language which could be included.

No criterion regarding quality of the papers in terms of publication in a peer-review journal was used, as there exists little evidence that there is any difference between different publication sources. However, when uncertainties in the reporting were encountered, this could lead to the exclusion of a paper, although this is really an application of the criteria of conversion and statistical details. All studies that had, in principle, studied the association between personality and traffic accident involvement, but were excluded for various reasons, are listed in Table 3 in af Wåhlberg, Barraclough and Freeman (submitted).

2.4 Studies excluded from the Clarke and Robertson data set

The criteria had the effect of excluding some of the studies included in the Clarke and Robertson meta-analysis. Out of thirty-three studies on traffic accidents included by Clarke and Robertson, sixteen were included in the present study.

The present analysis only considered studies that operationalized traffic accidents as the dependent variable, while the Clarke and Robertson study included both work and traffic accidents (but ran separate analyses). However, some of the papers which were listed as work-related in Table 1 of that study were actually about traffic accidents, although these were incurred as part of the subjects' work. These papers were included in the present analysis, with exceptions listed in af Wåhlberg, Barraclough and Freeman (submitted), Table 3.

2.5 Conversion of personality variables

The conversion of personality scales into Big Five dimensions were based upon empirically discovered correlations (see af Wåhlberg, Barraclough and Freeman, submitted, Table 4) or, if empirical associations were not available, the expert judgements from Salgado (2003). In that study, two researchers compared the definitions and items of different personality scales and

decided upon what Big Five dimension different personality inventory dimensions could be considered to be equal to.
For example, in a study by Evans and Rothbart (2007), it was found that the Negative affect scale of the Adult Temperament Questionnaire correlated most strongly with the N dimension of the Big Five. Therefore, effects for the Negative affect scale were noted under N in the data spreadsheet. In a few cases, ad hoc decisions were also made based upon the similarity of a scale with others. All conversions and data extractions were undertaken by the first author.

2.6 Conversion and calculation of effects in papers
The Pearson correlation was chosen as the common effect size statistic to use, as this was the most commonly reported one. Whenever possible, eligible papers with other types of statistics reported were converted into r, using the formulas by Wolf (1986). If more than one relevant value was reported (i.e. several scales could be converted into a BF dimension), these correlations were averaged by squaring each, summing them, dividing by number of correlations and taking the square root of the resulting value.
If a study only reported the significant findings (as for example Conger et al., 1959; Panek, Wagner, Barrett & Alexander, 1978), or only p-levels but no effect sizes, the largest possible effects, given the p-levels, were calculated. This method will over-estimate the effect, because smaller effects will be excluded or over-estimated. However, if it was not known how many scales had been used, the study was excluded, as the combination of over-estimation of single effects and exclusion of small ones was deemed to produce too high a level of bias.

2.7 Accident data and method effects
The accident data had two important features that were coded; the type of method used to gather it (self-report and/or archive) and the amount of variance (indicated by the mean and the standard deviation). As the hypothesis here was that studies using self-reports to predict the self-reported personality data would yield inflated effect sizes due to common method variance, these sources were separately coded. If a study had used both sources, an intermediate value was used.
Accident prediction studies all suffer from the problem of a highly skewed distribution and restriction of range of the dependent variable. More importantly, from a meta-analytic perspective, however, these problems differ between studies, and can be shown to affect the effect sizes (af Wåhlberg, 2009). These differences between samples are due to differences in the time periods for measuring accidents (af Wåhlberg, 2003), in risk between environments (e.g. Clarke & Robertson, 2005), and mileage (af Wåhlberg, 2009). Ideally, these three differences should be controlled for in meta-analysis. However, the only information which is regularly available is that of the time period for measurements and the mean of the accident variable. The interesting fact about this is that the latter reflects all these three factors; time period, risk and exposure. As could therefore be expected, the accident mean of the sample has been shown to be a very strong determinant of the effect size in the studies, usually outperforming time period, which is also strongly related to effect sizes (af Wåhlberg, 2009; af Wåhlberg, Barraclough & Freeman, 2015; Barraclough, af Wåhlberg, Freeman, Watson & Watson, 2016). Therefore, in the present study, the accident mean was controlled for in some calculations, with the aim to even out differences in time periods, risk and exposure between samples.
In some cases, the mean number of accidents was not reported, but the sample was divided into accident-involved and not accident-involved (e.g., basically all studies using t-tests and chi^2). In these cases, the numbers were calculated as one for the accident group drivers and zero for the controls. This would of course not be entirely true, but very few drivers have more than one crash in such time periods as those studied in this literature.

In all cases where it was possible, the effects included were for all accidents, i.e. not for sub-categories such as culpable or active. The all accidents variable is not the optimal criterion (af Wåhlberg, 2008; 2009), but the most commonly used one, and any other choice would have severely limited the number of available studies.

2.8 Identifying outlier effects

In a preliminary analysis, outlying values were identified, with the aim of excluding such values as probably faulty. If no explanation for the discrepancies could be found, such as obvious errors of transcription, these values were deleted. The definition of an outlier in the present work was based upon bi-variate associations, instead of uni-variate ones (see Ben-Gali, 2005). This was due to the fact that large differences were expected, mainly due to differences in accident variance between studies. A report which used a very high or very low mean of accidents could therefore yield an effect which would be deemed an outlier if considered in comparison to other effect sizes, but would be deemed a normal range-value if the variance in the accident variable was taken into account.

To identify outliers, the bi-variate associations for the factors versus the mean of accidents were calculated, and the Poom method was applied. The Euclidian distance from the regression line of each bi-variate point in a scatter plot was calculated, and these values treated as a variable. After calculating the mean of this variable (which is very close to zero) and standard deviation, the criterion of two standard deviations can be applied for identification of bi-variate outliers.

2.9 Variables extracted

For each study, all effects for personality versus accident record were coded. Furthermore, the mean number of accidents and the standard deviation of this variable, number of subjects, country of origin and details about the sample. Also, year of publication, type of source for the accident data (self-report or archive or both) and whether it used a Big Five instrument or not were coded.

2.10 Analysis

A random effects model was applied for all mean effects analyses (using the software Comprehensive Meta-Analysis), as method effects were expected. Also, an alternative calculation deploying the Hunter-Schmidt method (the mean effect of r weighed by N, as described by Field & Gillett, 2010) was performed, to check whether this could influence the results.

It can be noted that the confidence intervals for the random effects model only apply to this specific dataset, and those with similar accident means. This is due to the fact that there is really no such thing as a true population effect in these data. The effect size in accident prediction studies is mainly dependent upon the variation in the accident variable (af Wåhlberg, 2009; af Wåhlberg & Dorn, 2009; af Wåhlberg, Barraclough & Freeman, 2015; Barraclough, af Wåhlberg, Freeman, Watson & Watson, 2016). Therefore, any mean and confidence interval calculated are specific for a set of studies with this mean number of accidents as the dependent variable. Instead, calculations with accident mean held constant were used to control for such differences, and make the results comparable to those of other meta-analyses with different accident means (further described in the Discussion), thus making meta-analytic results of individual differences in accident record cumulative knowledge (Hedges, 1987). For this end, Pearson correlations were run between the Big Five dimensions and the accident mean to check whether the latter was indeed a moderator, in line with the recommendation by Steel and Kammeyer-Mueller (2002). As accident variance would (probably) not be equal between BF dimensions, effects were calculated with the mean

of accidents held constant, using the regression equations from the correlation of each dimension with the accident mean.
Results were not corrected for reliability of the personality and accident variables. This was due to the fact that a great many different scales were included, possibly with very different reliabilities, and correction would therefore have needed information on each scale. Also, it is uncertain whether the concept of reliability is applicable to traffic accidents.
To test for biases, correlations were run between effect size, year of publication and number of subjects, the first of which detects whether effects tend to change with time. This phenomenon can be due to changes in methodology, a change in dissemination bias over time, or an actual change in population effect size (e.g. Bond & Smith, 1996). All of these, however, indicate that there is no stable mean effect size. The effect size/N analysis detects whether effects are equally distributed on both sides of the mean. If not, this might indicate dissemination bias.
Moderator analyses were run using ordinary correlations for continuous variables and random effects models for the two categories (e.g. self-reported versus recorded accident data, and Big Five versus non-Big Five inventories).

Table 1: The correlations between effects for the BF dimensions, and with accident mean. Above the diagonal, the full dataset, under it the restricted set, after bivariate 2 std outlier deletions on BF/accident correlations. Along the diagonal, the correlations of the full and restricted datasets with the means of accidents.

BF dimension	E	A	C	N	O
Extraversion	**.098 N=49** **.122 N=46**	-.335 (N=28)	-.022 (N=28)	-.085 (N=35)	.140 (N=21)
Agreeableness	-.343 (N=24)	**-.104 N=26** **-.443* N=23**	.420* (N=23)	-.341 (N=27)	.186 (N=20)
Conscientiousness	.199 (N=26)	.249 (N=21)	**-.268 N=35** **-.395* N=32**	-.211 (N=25)	-.153 (N=20)
Neuroticism	.123 (N=33)	-.325 (N=25)	-.133 (N=22)	**.179 N=41** **.307 N=38**	-.438 (N=21)
Openness	-.205 (N=20)	.243 (N=19)	-.153 (N=20)	-.438 (N=21)	**-.374 N=19** **-.374 N=19**

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

3. Results

3.1 General

Besides the overall aim of summarising the research information on personality as a traffic accident predictor, the present study also aimed to report all the information used in the analysis, so as to facilitate future analysis by other researchers. Also, as many decisions, for example about conversion of scales into Big Five, can be criticised, it was deemed important to report this in full detail, so that any errors can be corrected by other researchers.
As the amount of information for each study was rather large, two complimentary tables were constructed, one containing mainly background information (Table a), and the other personality scale and effects data (Table b). Furtermore, data were ordered into three sections; those that were included in Clarke and Robertson (2005) (Tables 1ab, af Wåhlberg, Barraclough and Freeman, submitted) and in the present analysis, and those that were included only here (Tables 2ab, af Wåhlberg, Barraclough and Freeman (submitted). Finally, a table was constructed which listed all the papers that measured personality and traffic accidents, but which were excluded from the present analysis for various reasons (Table 3, af Wåhlberg, Barraclough and Freeman (submitted).
In total, sixty-two papers yielded sixty-eight different samples, and a further nine alternative calculations (e.g. for different time periods or accident data sources). These papers contained

a total of one hundred and ninety-two effect sizes which were included in the preliminary analysis (before cleansing from outliers).

3.2 Analysis and deletion of outliers

The Poom method was used to calculate bi-variate dispersion between each BF dimension effects and the mean of accidents, and all values outside of two standard deviations were removed (see Ben-Gali, 2005, for information about bi-variate outlier detection methods). It can be noted that this restricted dataset probably also contained outliers, if new Poom values had been calculated, but these were left in, as the standard convention for outlier deletion does not allow for repeated operations. Several samples did not have accident means reported, and possibly faulty values among these could not be identified. One outlying value was also deleted from the accident mean variable; 11.4 for the total time period in Achoui (2004).

To check whether the outlier detection had the intended effect of removing erroneous values, correlations were run among the BF dimensions and between these and the mean of accidents in samples, before and after exclusion of outliers (see Table 1). It can be seen that the effects in different dimensions tended to correlate strongly with each other and with the accident mean. For most of the BF inter-correlations, the accident mean explained the majority of the association, thus indicating that systematic variance between studies was largely due to variation in the accident mean.

3.3 Analysis of mean effects

First, the mean r weighted by N was calculated, by multiplying each effect by the number of subjects of the study, summing these values over studies and dividing this value with the total number of subjects in all studies. Second, random effects models were applied. These values were compared with those of Clarke and Robertson (Table 2). There was fair agreement between the two different types of statistical methods as well as between calculations with and without outliers, meaning that all the effects in the present analysis were much smaller than those of Clarke and Robertson.

In the last column in Table 2, the regression equations from Table 1 correlations have been used to calculate the expected effect in a study with a sample having an accident mean of 1. This is similar to correcting for restricted variance, but is empirically (based upon data), instead of statistically (based upon mathematical assumptions), based (see Hurtz & Donovan, 2000, for a similar method). This method is further explained in the Appendix.

3.4 Heterogeneity

Heterogeneity (whether there is variation between samples which cannot be explained by sampling error) was calculated for the restricted samples. For this purpose, the Q and I^2 statistics were used. The first indicates whether there is a significant degree of heterogeneity in the sample, while the second quantifies this amount, i.e. it indicates the percentage of variation between studies that is not due to sampling error. It is independent of the number of samples, while the Q statistic is dependent upon both the actual variation and the number of studies when significance is calculated (Higgins & Thompson, 2002; Huedo-Medina, Sanchez-Meca, Marin-Martinez & Botella, 2006). Table 3 depicts the results for the heterogeneity analysis. Both E and C yielded rather large amounts of unexplained variance, despite deletion of outliers.

Table 2: Mean effects over studies, calculated in different ways. First, results from Clarke and Robertson (2005)(sign change for C and A), second, from the present study. Shown are the number of samples (k), total number of subjects (N) and mean effects for each Big Five dimensions, mean r weighted for N, and calculated in a random effects model. Finally, a calculation of what r would be if the accident mean was 1, calculated from the regression formulas for the correlations between effects and mean of accidents.

	C & R results (traffic only)			Present study, r0				Present study r1				
Dimension	k	Total N	Mean r	k	Total N	Mean r weighted for N	Mean r (CI) in a random model	k	Total N	Mean r weighted for N	Mean r (CI) in a random model	Estimated r when accident mean=1 #
Extraversion	16	4424	.146	57	43884	.053	.064 (0.46/.083)	54*	43751	.053	.065 (.047/.082)	.083 (k=46)
Agreeableness	7	3108	-.127	29	10577	-.072	-.072 (-.049/-.095)	26**	10161	-.071	-.071 (-.051/-.090)	-.083 (k=23)
Conscientiousness	9	3425	-.160	36	23873	-.065	-.082 (-.056/-.108)	33***	23644	-.062	-.071 (-.048/-.094)	-.096 (k=32)
Neuroticism	8	1460	.062	49	23452	.023	.038 (.019/.057)	46****	23337	.023	.036 (.018/.054)	.106 (k=37)
Openness	3	577	.077	21	8190	.011	.016 (-.014/.045)	21	8190	.011	.016 (-.014/.045)	.005 (k=19)

Study values excluded from r1
* Carty, Stough and Gillespie, 1998; Pestonjee & Singh, 1980; Andersson, Nilsson and Henriksson, 1970
** Roy and Choudhary, 1985; Yates et al., 1987; Arthur & Day, 2008, Study 1
*** West, Elander and French, 1993; Arthur and Doverspike, 2001; Schwebel et al., 2006
**** Plummer & Sunder Das, 1973; Evans, Palsane and Carrere, 1987; Achoui, 2004 (per year)
# In this column, k refers to the number of studies used in the regression equation for the personality variable versus accidents, from which the expected r when accidents=1 was calculated.

Table 3: The heterogeneity of effects in the restricted samples (r1). Shown are the number of studies, and the Q and I squared statistics. Significance set at p<.05.

	k	Q	I square
Extraversion	54	102.3 (p<.001)	48.2
Agreeableness	26	21.0 ns	0
Conscientiousness	33	65.1 (p<.001)	50.8
Neuroticism	46	53.1 ns	15.2
Openness	21	31.1 ns	35.7

3.5 Effect bias due to N and time

Correlations were run between effect sizes (r^2), year of publication and the number of subjects. With an initial negative sign, as for the A and C dimensions, decreasing effect sizes yields a positive correlation with year of publication. As can be seen in Table 4, effects tended to decrease (although not always significantly so) over time for all dimensions but O. However, as the mean effect was in the contrary direction to the theoretical expectation, this increase was not a positive thing.

Effects for E had the strongest association with number of subjects. Apparently, a few small studies had yielded very high values. It is notable that the two studies reporting the largest effects (Hartman & Rawson, 1992; Loo, 1978) did not report the mean for their accident variable, and were therefore not included in the outlier cleansing operation.

Duval and Tweedie's trim-and-fill operation yielded results similar to the correlation between N and effects, in that the E and N dimensions were found to have slightly skewed distributions. Only for N was there a substantial reduction of the mean effect size.

It can be noted that the number of subjects tended to increase over time (r=.247, k=77, p<.05). The reduction of effect sizes over time could therefore be due to smaller effects being significant, and therfore published, in later studies.

Table 4: Publication bias calculations. The correlations between effects (r^2) and the year the study was published and the number of subjects. Also, the number of added studies in a trim-and-fill analysis, and the adjusted mean value, if any.

Variable	Extraversion (k=54)	Agreeableness (k=26)	Conscientiousness (k=33)	Neuroticism (k=46)	Openness (k=21)
Year of publication	-.302*	.393*	.161	-.057	.378
N of study	-.164	.068	.118	-.148	.036
Trim-and-fill	8 (.053)	0	0	16 (.017)	0

* p<.05

3.6 Analysis of method effects

First, differences between accident sources were calculated (Table 5). Here, a number of effects were deleted, because they had used a combination of self-reports and records. In all comparisons, CIs overlapped. In the case of E, records actually yielded the larger effect. Second, studies that used Big Five instruments, and others, were compared. In all five cases, the BF instruments yielded much smaller effects, but only in the case of O did CIs not overlap. This difference between instruments probably caused most of the effect of year of publication, as the Big Five studies were on average published almost twelve years later.

Table 5: The differences in effect size between studies categorized as to whether they used self-reported or recorded accidents, and Big Five instruments or not, in the r1 samples. Random effects models applied. Studies using a combination of self-reports and records were deleted from the data source calculations (E=3, A=1, C=1, N=3).

	Self-reported accidents		Recorded accidents		Big Five personality traits		Other inventories	
Dimension	k	Mean r (CI)	k	Mean r (CI)	k	Mean r (CI)	k	Mean r (CI)
Extraversion	39	.057 (.039/.076)	12	.104 (.045/.162)	18	.059 (.030/.088)	36	.069 (.045/.092)
Agreeableness	20	-.074 (-.053/-.095)	5	-.042 (.013/-.097)	18	-.068 (-.046/-.089)	8	-.083 (-.039/-.128)
Conscientiousness	27	-.077 (-.052/-.103)	5	-.023 (.031/-.077)	19	-.073 (-.041/-.105)	14	-.067 (-.030/-.104)
Neuroticism	32	.045 (.021/.070)	11	.012 (-.006/.029)	20	.025 (.002/.049)	26	.056 (.026/.086)
Openness	16	.026 (.002/.051)	4	-.022 (.072/-.115)	18	.029 (.005/.052)	3	-.097 (-.040/-.154)

3.7 Accident mean and standard deviation

The correlation between the mean and the standard deviation of the accident variable was .91 (N=44, p<.001). This shows that the mean of the accident variable is a viable proxy for the variance in the variable. However, comparing the correlations with effect sizes for these two variables yielded no clear-cut advantage for either of them. Theoretically, the standard deviation should be superior.

4. Discussion

4.1 Results

This paper aimed to investigate the association between personality, as expressed in the Big Five dimensions, and traffic accident involvement. Although the objective was to estimate the population effect, most of the calculations were undertaken to check and control for possible moderators, including variance in the accident variable, source of the data and dissemination bias. After controlling for such effects, the population estimates were found to be in the range of .01-.07. How does this compare to other predictors of accident involvement?

There are other meta-analytic results for accident prediction which can be compared to the present ones, which indicate that even when restriction of variance is held constant, personality is not a good predictor of accident involvement. If accidents are predicted from previous accident involvement, and the predicted period has a mean of 1 accidents in the sample (as in the calculation in the last column of Table 2, see the Appendix for a detailed explanation), the correlation is likely to exceed .2, which is more than the fourfold value of any of the BF dimensions (calculation based upon the meta-analysis of stability of accident record over time in af Wåhlberg, 2009). Similarly, traffic offences have a weighted mean effect of .18 (Pearson r, CI .17/.20), which is much larger than those for personality, despite the offences variable also suffering from restricted variance. When the accident mean is held equal to 1, the expected effect is .25 (Barraclough, af Wåhlberg, Freeman, Watson & Watson, 2016), as compared to a maximum of .106 for (a single dimension of) personality.

Also, Arthur, Barrett and Alexander (1991) reported a mean correlation of .257 for selective attention and .151 for perceptual style, along with a number of other results. On the other hand, the effects of personality are similar to those for a diagnosis of ADHD, which yielded mean relative risks of 1.2-1.3 in Vaa (2014), and complex reaction time (r=.053) and level of distress (r=.023) in Arthur, Barrett and Alexander (1991). Although these latter results did not control for differences in accident variance, they still give some indication of how well personality (as measured by self-reports) do in comparison to other predictors.

It should be evident that accident involvement can be predicted with a fair degree of precision if variance in the accident variable is high (approaching a normal distribution), but that not all variables are equally strong predictors. However, the methodology of most studies on accident prediction is sub-optimal (af Wåhlberg, 2009), and effects are therefore usually under-estimated in each study. Therefore, the current results can be improved upon, i.e. larger effect sizes can probably be found by using better methodology.

The current meta-analysis, on the other hand, has probably over-estimated the correlation between self-reported personality and traffic accidents in terms of published evidence. This is due to several factors; the inclusion of a few studies with only p-values reported, the exclusion of several studies that reported only that the results were not significant (e.g. Häkkinen, 1979), and the declining trend of effects over time. It is therefore predicted that a new meta-analysis in a few years time, using a stricter criterion concerning statistical details, will yield even (slightly) smaller mean effects, if the current methodology continues to be used.

On the other hand, it was also apparent that the effects were strongly dependent upon the variance in the accident variable. Studies using high-risk samples over long time periods will therefore yield larger effects. The 'true' correlation between self-reports of personality and accidents is therefore larger, in similarity to what can be found if variables are corrected for restricted variance. It is predicted that studies using high-risk samples for long time periods, yielding accidents means >1 will have larger effects than those in the published evidence up to now (af Wåhlberg, 2009).

No reliable evidence of common method variance could be detected in the present data. However, as none of the personality dimensions explain even one percent of the variance in accidents, there is not really much of an effect which could be caused by common method variance anyway. If there is an effect of common method variance in personality inventories when applied to traffic safety, this effect is apparently nullified by the poor validity of the self-reports.

In the present study, there was no tendency for the Big Five inventories to have larger effects than other personality taxonomies, such as the Eysenck Personality Questionnaire, but rather the opposite (four out of five dimensions yielding lower estimates). This is similar to what has been found for job performance indicators (Hurtz & Donovan, 2000).

Another central finding of this analysis was that the quality of research on personality as a predictor of traffic accident involvement is rather poor, in terms of yielding cumulative knowledge. For example, a large number of studies which had gathered relevant data had simply not reported the statistics needed for a meta-analysis. This is not something which is peculiar for this area of research (Wagenaar, Zobeck, Williams & Hingson, 1995, reported even more severe problems), but advances in road safety knowledge would be enhanced if key information is included in published findings (see af Wåhlberg, 2010, for a traffic safety-specific guide on what information should be included). Briefly, this includes the time period, mean and standard deviation of the accident variable, the source of the data, zero-order correlations between all variables, and how exposure and culpability have been handled.

These results are not surprising, as self-reports in general are dubious as to their validity (see af Wåhlberg, 2009, for a traffic-psychology specific summary). As for self-reports of personality, Costa and McCrae (1988) found correlations between self-reported and spouse-reported personality on the NEO-PI to be around .5. Mount, Barrick and Perkins Strauss (1994) reported lower validities of self-reports as compared to observer ratings versus several job performance, and also reviewed several papers which indicated the same result. The same effect was reported from a meta-analysis by Connelly and Ones (2010). As a consequence of this, the predictive capacity of personality in terms of repeated behaviours, not questionnaire responses, versus crashes, is still largely unknown.

4.2 Limitations

The main limitations of the present analysis, in terms of how predictive self-reports of personality are of traffic accidents, are the papers included and how effects were converted. It can be claimed that a) some studies should not have been included at all, because the measures are not of personality, b) some studies have been erroneously excluded, c) some measures have been erroneously converted into the wrong BF dimension. However, as the total number of subjects in the present analysis was rather respectable, and most conversions were rather straightforward (e.g.,i.e. studies have shown a fairly strong correlation with one BF dimension only), and the studies using BF had about twenty-one percent of the total number of subjects, no large differences are likely.

The current study used the Big Five as an organizing framework, and it is acknowledged that the results and conclusions are therefore only valid for this system. As pointed out by a reviewer, the Big Five are rather general variables, and stronger results could possibly be achieved by more narrowly defined personality traits (facets).

In the end, the present study indicates that although there is a wealth of papers available on the association of personality and accident involvement, this research is rather disparate. This fact strongly influences what conclusions can be drawn from a meta-analysis, as the choices made about what to include largely determines the results. Such a state of the art therefore leads to the need amongst traffic safety researchers to discuss what kinds of methods are actually reliable, and to test them. Such discussions and tests are currently rather unusual (af Wåhlberg, 2009) but are clearly warranted if researchers are collectively going to contribute to reducing the personal, social and economic burden of road crashes.

4.3 Conclusions

Taken together, the present meta-analysis has found personality, in terms of single occasion self-reported questionnaire responses, to have little predictive value for traffic accident involvement, in agreement with the results of Salgado (2002) for workplace accidents.

It is therefore recommended that future research focus on three features which might yield larger effects. First, testing more specific personality traits (such as sensation seeking), as the broad factors of the Big Five apparently do not work well as predictors of crash involvement. Second, measuring personality using methods other than self-report (as recommended by Connolly & Ones, 2010). Third, testing whether repeated self-reports of personality can increase the effects (as this method tends to increase the reliability and validity of measurements by removing random error). These methods could possibly yield larger effect sizes and a truer picture concerning the effects of personality on traffic accident involvement.

Acknowledgements: A number of researchers have responded to e-mailed questions about their data, sent papers and/or further statistical results, and the authors would therefore like to thank the following colleagues; Mark Conner (University of Leeds, UK), Pauline Gulliver (New Zealand), Joan Harvey (Newcastle University, UK), Dragan Jovanovic (Department of Transport, Serbia), Timo Lajunen (Middle East Technical University, Turkey), Inese Muzikante (University of Latvia, Latvia), Laura Seibokaite (Vytautas Magnus University, Lithuania), Nebi Sümer (Middle East Technical University, Turkey), Heikki Summala (University of Helsinki, Finland), Orit Taubman-Ben-Ari (Bar Ilan University, Israel), and Lisa Wundersitz (University of Adelaide, Australia).

The work of Anders af Wåhlberg was supported by a scholarship from the Ax:son Johnson Foundation (Sweden), while Peter Barraclough worked on this as part of his PhD, funded by the Australian Research Council Discovery Grant DP130101443.

5. References

Achoui, M. (2004). *Personality and Driving Accidents.* Unpublished manuscript.

Adams, J. R. (1970). Personality variables associated with traffic accidents. *Behavioral Research in Highway Safety, 1,* 3-18.

Aguinis, H., Pierce, C. A., Bosco, F. A., Dalton, D. R., & Dalton, C. M. (2011). Debunking myths and urban legends about meta-analysis. *Organizational Research Methods, 14,* 306-331.

Andersson, A. L., Nilsson, A., & Henriksson, N. (1970). Personality differences between accident-loaded and accident-free young car drivers. *British Journal of Psychology, 61,* 409-421.

Arthur, W. Jr., Barrett, G. V., & Alexander, R. A. (1991). Prediction of vehicular accident involvement: A meta-analysis. *Human Performance, 4,* 89-105.

Arthur, W. Jr., & Day, E. A. (2008). Information processing, personality, and demographic variables as predictors of crashes and moving violations. In G. P. Bartley (Ed.) *Traffic Accidents: Causes and Outcomes.* Nova.

Arthur, W., Jr., & Doverspike, D. (2001). Predicting motor vehicle crash involvement from a personality measure and a driving knowledge test. *Journal of Prevention and Intervention in the Community, 22,* 35-42.

Arthur, W., Jr., Tubre, T. C., Day, E., Sheehan, M. K., Sanchez-Ku, M. L., Paul, D. S., Paulus, L. E., & Archuleta, K. D. (2001). Motor vehicle crash involvement and moving violations: Convergence of self-report and archival data. *Human Factors, 43,* 1-11.

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

Bax, L., & Moons, K. G. (2011). Beyond publication bias. *Journal of Clinical Epidemiology, 64*, 459-462.

Begg, D. J., Langley, J. D., & Williams, S. M. (1999). A longitudinal study of lifestyle factors as predictors of injuries and crashes among young adults. *Accident Analysis and Prevention, 31,* 1-11.

Beirness, D. J. (1993). Do we really drive as we live? The role of personality factors in road crashes. *Alcohol, Drugs and Driving, 9,* 129-143.

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

Bond, R., & Smith, P. B. (1996). Culture and conformity: A meta-analysis of studies using Asch's (1952b, 1956) line judgement task. *Psychological Bulletin, 119,* 111-137.

Brandau, H., Daghofer, F., Hofmann, M., & Spitzer, P. (2011). Personality subtypes of young moped drivers, their relationship to risk-taking behavior and involvement in road crashes in an Austrian sample. *Accident Analysis and Prevention, 43,* 1713-1719.

Burns, P. C., & Wilde, G. J. (1995). Risk taking in male taxi drivers: Relationships among personality, observational data and driver records. *Personality and Individual Differences, 18,* 267-278.

Carty, M., Stough, C., & Gillespie, N. (1998). The psychological predictors of work accidents and driving convictions in the transport industry. *Safety Science Monitor, 3,* 1-13.

Chung, K. C., Burns, P. B., & Kim, H. M. (2006). A practical guide to meta-analysis. *Journal of Hand Surgery, 31A,* 1671-1678.

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. *Journal of Occupational and Organizational Psychology, 78,* 355-376.

Clement, R., & Jonah, B. A. (1984). Field dependence, sensation seeking and driving behaviour. *Personality and Individual Differences, 5,* 87-93.

Conger, J. J., Gaskill, D. D., Glad, L., Hassel, L., Rainey, R. V., Sawrey, W. L., & Turrell, E. S. (1959). Psychological and psychophysiological factors in motor vehicle accidents. *Journal of the American Medical Association, 169,* 1581-1587.

Connelly, B. S., & Ones, D. S. (2010). An other perspective on personality: Meta-analytic integration of observer's accuracy and predictive validity. *Psychological Bulletin, 136,* 1092-1122.

Costa, P. T., & McCrae, R. R. (1988). Personality in adulthood: a six-year longitudinal study of self-reports and spouse ratings on the NEO Personality Inventory. *Journal of Personality and Social Psychology, 54,* 853-863.

Donovan, D. M., Marlatt, G. A., & Salzberg, P. M. (1983). Drinking behaviour, personality factors and high-risk driving. *Journal of Studies on Alcohol, 44,* 395-428.

Drake, C. A. (1940). Accident proneness: A hypothesis. *Character and Personality, 8,* 335-341.

Eagly, A. H., Makhijani, M. G., & Klonsky, B. G. (1992). Gender and the evaluation of leaders: A meta-analysis. *Psychological Bulletin, 111,* 3-22.

Evans, D. E., & Rothbart, M. K. (2007). Developing a model for adult temperament. *Journal of Research in Personality, 41,* 868-888.

Evans, G. W., Palsane, M. N., & Carrere, S. (1987). Type A behavior and occupational stress: A cross-cultural study of blue-collar workers. *Journal of Personality and Social Psychology, 52,* 1002-1007.

Eyding, D., Lelgemann, M., Grouven, U., Härter, M., Kromp, M., Kaiser, T., Kerekes, M. F., Gerken, M., & Wieseler, B. (2010). Reboxetine for acute treatment of major depression: systematic review and meta-analysis of published and unpublished placebo and selective serotonin reuptake inhibitor controlled trials. *BMJ, 341,* c4737. doi:10.1136/bmj.c4737

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

Fournier, K. A., Hass, C. J., Naik, S. K., Lodha, N., & Cauraugh, J. H. (2010). Motor coordination in autism spectrum disorders: A synthesis and meta-analysis. *Journal of Autism Developmental Disorder, 40,* 1227-1240.

Greenwood, M., & Woods, H. M. (1919). *The Incidence of Industrial Accidents upon Individuals with Specific Reference to Multiple Accidents.* Industrial Fatigue Research Board, Report No 4. London: His Majesty's Stationary Office.

Groh, A. M., Fearon, R. P., Bakermans-Kranenburg, M. J., van Ijzendoorn, M. H., Steele, R. D., & Roisman, G. I. (2014). The significance of attachment security for children's social competence with peers: a meta-analytic study. *Attachment & Human Development, 16,* 103-136.

Hansen, C. P. (1988). Personality characteristics of the accident involved employee. *Journal of Business and Psychology, 2,* 346-365.

Harris, F. J. (1950). Can personality tests identify accident-prone employees? *Personnel Psychology, 3,* 455-459.

Hartman, M. L., & Rawson, H. E. (1992). Differences in and correlates of sensation seeking in male and female athletes and nonathletes. *Personality and Individual Differences, 13,* 805-812.

Hedges, L. V. (1987). How hard is hard science, how soft is soft science? The empirical cumulativeness of research. *American Psychologist, 42,* 443-455.

Hessing, D. J., Elffers, H., & Weigel, R. H. (1988). Exploring the limits of self-reports and Reasoned Action: An investigation of the psychology of tax evasion behavior. *Journal of Personality and Social Psychology, 54,* 405-413.

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

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.

Huf, W., Kalcher, K., Pail, G., Friedrich, M.-E., Filzmoser, P., & Kasper, S. (2011). Meta-analysis: Fact or fiction? How to interpret meta-analyses. *The World Journal of Biological Psychiatry, 12,* 188-200.

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

Hurtz, G. M., & Donovan, J. J. (2000). Personality and job performance: The Big Five revisited. *Journal of Applied Psychology, 85,* 869-879.

Häkkinen, S. (1979). Traffic accidents and professional driver characteristics: A follow-up study. *Accident Analysis and Prevention, 11,* 7-18.

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.

Jamison, K., & McGlothlin, W. H. (1973). Drug usage, personality, attitudinal, and behavioral correlates of driving behavior. *The Journal of Psychology, 83,* 123-130.

Jonah, B. A. (1997). Sensation seeking and risky driving: A review and synthesis of the literature. *Accident Analysis and Prevention, 29,* 651-665.

Judge, T. A., Colbert, A. E., & Ilies, R. (2004). A meta-analysis of the relationship between intelligence and leadership. *Journal of Applied Psychology, 89,* 542-552.

Lawton, R., & Parker, D. (1998). Individual differences in accident liability: A review and integrative approach. *Human Factors, 40,* 655-671.

Lester, J. (1991). *Individual Differences in Accident Liability: Review of the Literature.* TRRL Research Report 306. Crowthorne: Transport and Road Research Laboratory.

Loo, R. (1978). Individual differences and the perception of traffic signs. *Human Factors, 20,* 65-74.

Macaskill, P., Walter, S. D., & Irwig, L. (2001). A comparison of methods to detect publication bias in meta-analysis. *Statistics in Medicine, 20,* 641-654.

McGuire, F. L. (1976). Personality factors in highway accidents. *Human Factors, 18,* 433-442.

Morina, N., Ijntema, H., Meyerbröker, K., & Emmelkamp, P. M. (2015). Can virtual reality exposure therapy gains be generalized to real life? A meta-analysis of studies applying behavioral assessments. *Behavior Research and Therapy, 74*, 18-24.

Mount, M. K., Barrick, M. R., & Perkins Strauss, J. (1994). Validity of observer ratings of the Big Five personality factors. *Journal of Applied Psychology, 79,* 272-280.

Møller, A. P., & Jennions, M. D. (2001). Testing and adjusting for publication bias. *Trends in Ecology & Evolution, 16,* 580-586.

Nichols, A. L., Classen, S., McPeek, R., & Breiner, J. (2012). Does personality predict driving performance in middle and old age? An evidence-based literature review. *Traffic Injury Prevention, 13,* 133-143.

Noble, J. H. (2006). Meta-analysis: Methods, strengths, weaknesses, and political uses. *Journal of Laboratory and Clinical Medicine, 147,* 7-20.

Orme-Johnson, D. W., & Dillbeck, M. . (2014). Methodological concerns for meta-analyses of meditation: Comment on Sedlmeier et al. (2012). *Psychological Bulletin, 140,* 610-616.

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

Parker, J. W. (1953). Psychological and personal history data related to accident records of commercial truck drivers. *Journal of Applied Psychology, 37,* 317-320.

Pestonjee, D. M., & Singh, U. B. (1980). Neuroticism-extraversion as correlates of accident occurence. *Accident Analysis and Prevention, 12,* 201-204.

Pham, B., Platt, R., McAuley, L., Klassen, T. P., & Moher, D. (2001). Is there a "best" way to detect and minimize publication bias? An empirical evaluation. *Evaluation and the Health Professions, 24,* 109-125.

Plummer, L. S., & Sunder Das, S. S. (1973). A study of dichotomous thought processes in accident-prone drivers. *British Journal of Psychiatry, 122,* 289-294.

Podsakoff, P. M., Mackenzie, S. B., Lee, J. Y., & Podsakoff, N. P. (2003). Common method biases in behavioral research: A critical review of the literature and recommended remedies. *Journal of Applied Psychology, 88,* 879-903.

Quenault, S. W. (1967). *Driver Behaviour - Safe and Unsafe Drivers.* RRL Report LR 70. Crowthorne: Road Research Laboratory.

Reijntjes, A., Kamphuis, J. H., Prinzie, P., & Telch, M. J. (2010). Peer victimization and internalizing problems in children: A meta-analysis of longitudinal studies. *Child Abuse & Neglect*, *34,* 244-252.

Roy, G. S., & Choudhary, R. K. (1985). Driver control as a factor in road safety. *Asian Journal of Psychology and Education, 16,* 33-37.

Salgado, J. F. (2002). The Big Five personality dimensions and counterproductive behaviors. *International Journal of Selection and Assessment, 10,* 117-125.

Salgado, J. F. (2003). Predicting job performance using FFM and non-FFM personality measures. *Journal of Occupational and Organizational Psychology, 76,* 323-346.

Schwartz, N. (1999). Self-reports: How the questions shape the answers. *American Psychologist, 54,* 93-105.

Schwebel, D. C., Severson, J., Ball, K. K., & Rizzo, M. (2006). Individual differences factors in risky driving: The roles of anger/hostility, conscientiousness, and sensation seeking. *Accident Analysis and Prevention, 38,* 801-810.

Shaw, L. (1965). The practical use of projective personality tests as accident predictors. *Traffic Safety Research Review, 9,* 34-72.

Shaw, L., & Sichel, H. S. (1971). *Accident Proneness.* Oxford: Pergamon.

Signori, E. I., & Bowman, R. G. (1974). On the study of personality factors in research on driving behavior. *Perceptual and Motor Skills, 38,* 1067-1076.

Song, F., Parekh, S., Hooper, L., Loke, Y. K., Ryder, J., & Sutton, A. J. (2010). Dissemination and publication of research findings: an updated review of related biases. *Health Technology Assessment, 14,* 1-193.

Steel, P. D., & Kammeyer-Mueller, J. D. (2002). Comparing meta-analytic moderator estimation techniques under realistic conditions. *Journal of Applied Psychology, 87*, 96-111.

Sterne, J. A., Gavaghan, D., & Egger, M. (2000). Publication and related bias in meta-analysis: power of statistical tests and prevalence in the literature. *Journal of Clinical Epidemiology, 53,* 1119-1129.

Vaa, T. (2014). ADHD and relative risk of accidents in road traffic: A meta-analysis. *Accident Analysis and Prevention, 62,* 415-425.

Vevea, J. L., & Woods, C. M. (2005). Publication bias in research synthesis: Sensitivity analysis using a priori weight functions. *Psychological Methods, 10,* 428-443.

Wagenaar, A. C., Zobeck, T. S., Williams, G. D., & Hingson, R. (1995). Methods used in studies of drink-drive control efforts: A meta-analysis of the literature from 1960 to 1991. *Accident Analysis and Prevention, 27,* 307-316.

West, R. J., Elander, J., & French, D. J. (1993). Mild social deviance, Type-A behaviour pattern and decision-making style as predictors of self-reported driving style and traffic accident risk. *British Journal of Psychology, 84,* 207-219.

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

Wolf, F. M. (1986). *Meta-analysis : Quantitative Methods for Research Synthesis.* Beverly Hills: Sage.

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. (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. (2010). A reporting guide for studies on individual differences in safety. *Journal of Safety Research, 41,* 381-383.

af Wåhlberg, A. E., Barraclough, P., & Freeman, J. (2015). The Driver Behaviour Questionnaire as accident predictor; a methodological re-meta-analysis. *Journal of Safety Research, 55,* 185-212.
DOI: 10.1016/j.jsr.2015.08.003

af Wåhlberg, A. E., Barraclough, P., & Freeman, J. (submitted). Meta-analytic data for personality as traffic accident predictor, and conversions between different personality systems. *Data in Brief,*

af Wåhlberg, A. E., & Dorn, L. (2009). Bus driver accident record; the return of accident proneness. *Theoretical Issues in Ergonomics Science, 10,* 77-91.

Yates, W. R., Noyes, R. Jr., Petty, F., Brown, K., & O'Gorman, T. (1987). Factors associated with motor vehicle accidents among male alcoholics. *Journal of Studies on Alcohol, 48,* 586-590.

Appendix: Calculating an expected effect size at accident mean=1

Most published studies on prediction of individual differences in traffic accident record suffers from restriction of range effects in the accident variable. This might be most obvious in studies which have dichotomized the accident variable, but there are usually very few subjects with more than one crash in a sample, unless it is a high-risk sample (professional drivers, for example), or the time period used is very long (>10 years). Basically, there are very few studies where the accident variable has a normal distribution.
This restriction of range differs between studies, because they have different accident means. This effect was first found in accident proneness studies, where accident record is predicted from previous accidents. In af Wåhlberg (2009), it was found that 70-80 percent of the variance in effect size between such studies could be explained by the accident mean. This is shown in Figure 1.
From this figure, it can by simple visual inspection be seen that there exist a very close association between the accident mean in a sample and the correlation between the accidents in two consecutive time periods (with the exception of one outlier). Furthermore, the regression equation for this correlation between mean of accidents and effect size indicates how much the effect increases with the level of accidents in the mean. Therefore, if accident mean is set at 1, the predicted effect will be .02781+.02531=.05312 percent explained variance, i.e. r=.23. From this equation, the expected correlation in a certain study can therefore be predicted with very good accuracy (and even better if the outlier is deleted), if the mean of accidents is known.

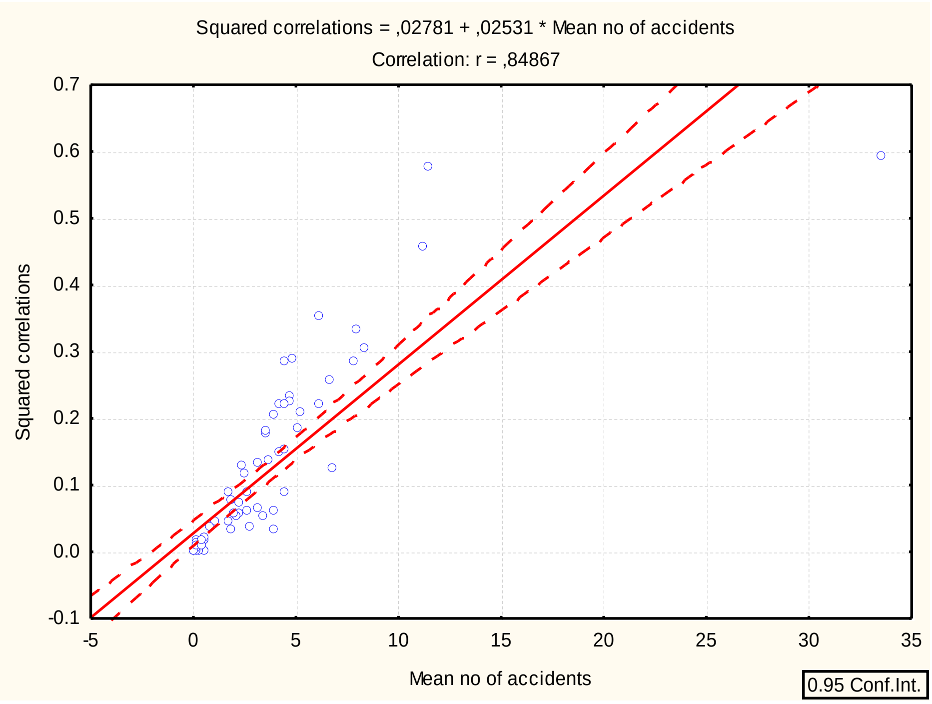

Figure 1: The association between mean number of accidents in each sample for the whole time period used and the (squared) correlation of accidents between parts of this period (usually split-half). N=78. Data from af Wåhlberg (2009).

However, this regression equation will be different for a different predictor, as it is expected that each predictor will have a different true population effect, which is the correlation when accidents are normally distributed. Therefore, a meta-analytically derived formula must be found for each predictor if an expected effect size at a certain level of accident mean is to calculated.

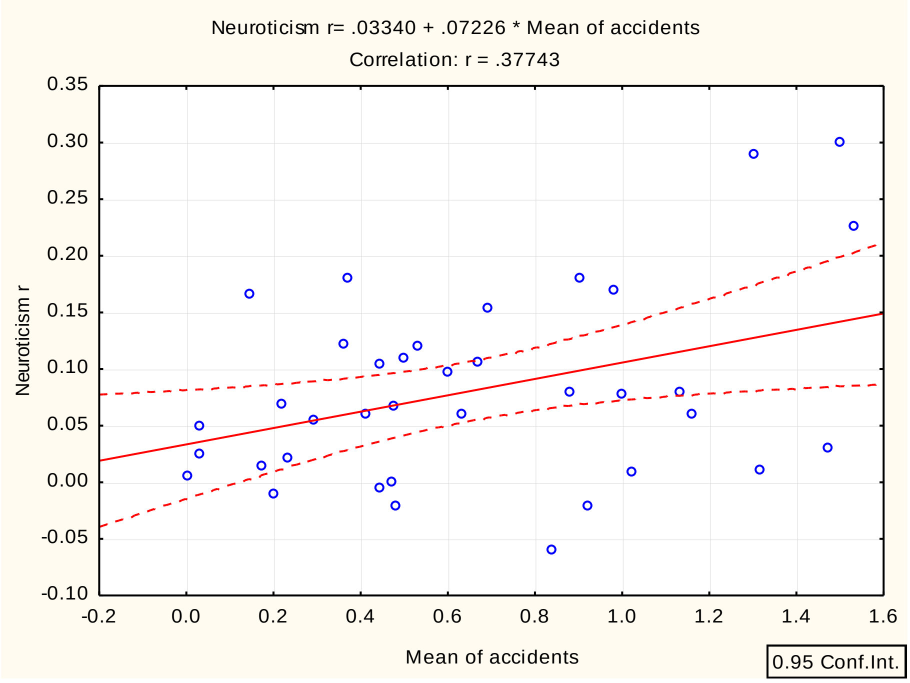

Figure 2: The association between mean number of accidents in each sample and the correlation between Neuroticism and accidents. N=37 (one accident outlier deleted).

In Figure 2, data from the present paper is used. Some differences between the two figures can be noted. First, Figure 1 used squared correlations, while Figure 2 used the raw correlations. The difference between these two types of calculations is very small as long as zero-order correlations are small. Second, the expected correlation in Figure 2 at accident mean=1 is, as reported in Table 2, is .106, much lower than that for accident-accident prediction in Figure 1. Third, the precision of the expected/predicted value in Figure 2 is much lower than in Figure 1, as the correlation is lower. It can also be inferred from these examples that the true population correlation (which happens when the accident variable is normally distributed) is higher for accident-accident prediction than for personality-accident.