. Pons de Villanueva J, Seguí-Gómez M. Validation of self-reported motor-vehicle crash and related work leave in a multipurpose prospective cohort. *Int J Inj Control and Safety Prom* 2010;12:1-8. PMID 20397084.

Validation of self-reported motor-vehicle crash and related work leave in a multi-purpose prospective cohort

Juan Pons-Villanueva[a] and María Seguí-Gómez[b]*

[a]*Preventive Medicine and Public Health Department, School of Medicine, University of Navarra, Pamplona, Spain;* [b]*Preventive Medicine and Public Health Department, European Center for Injury Prevention, School of Medicine, University of Navarra, Pamplona, Spain*

(Received 24 October 2009; final version received 27 January 2010)

Validation studies about self-reported motor vehicle crashes and related work leave are scarce. The Seguimiento Universidad de Navarra (SUN) is a multi-purpose cohort study undertaken in Spain. Motor vehicle crash risk factors are assessed within it. The objective of this study was to validate the incidents of self-reported motor vehicle crashes (MVC) and related work leave through mailing and clinical notes' review. The method included questions in the cohort's questionnaire regarding motor vehicle crash incident and work leave. We made both a re-test and a criterion validity study for self-reported answers. The results show a moderate κ Cohen's correlation coefficient for both events. Re-test reliability κ for MVC was 0.55, for MVC-related work leave it was 0.53. Criterion validity κ was 0.25 for MVC and 0.25 for MVC-related work leave. These results show a moderate agreement for re-test both for MVC and MVC-related work leave. For criterion validity, the results show a fair agreement. The magnitude of the agreement is similar to other similar studies and allows the use of these data in epidemiological studies.

Keywords: motor vehicle crash; validation; reliability; agreement; injury

1. Introduction

Among the multiple purposes of a cohort study in the Seguimiento Universidad de Navarra (SUN), Spain, is the assessment of risk factors associated with the incidence of motor vehicle-related injuries. This study is managed from Pamplona, Spain and its participants are university graduates from all over Spain. Recruitment and follow-up are made through mailed questionnaires. The open enrolment survey started in year 1999. The baseline questionnaire (Q_0) included amongst others, questions on motor vehicle crashes (MVC), as do all follow-up biennial questionnaires (Seguí-Gómez, de la Fuente, Vázquez, de Irala, & Martínez-González, 2006).

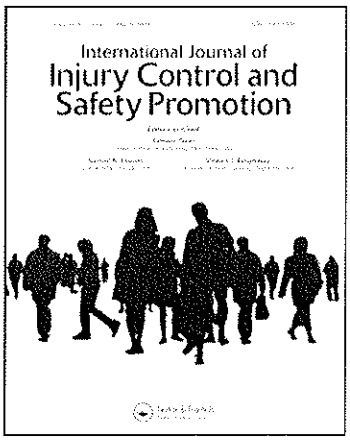
Validation of other event-related questions such as hypertension and physical activity has already been done and the results are encouraging (Alonso, Beunza, Delgado-Rodríguez, & Martínez-González, 2005; Martínez-González, López-Fontana, Varo, Sánchez-Villegas, & Martínez, 2005), as are validation exercises of other questions in other cohort studies (Barr, Herbstman, Speizer, & Camargo, 2002; Hu et al., 1999; Salvini et al., 1989; Willett et al., 2001). Yet, there are few studies assessing the reliability of self-reported information in motor vehicle-related injuries (Arthur et al., 2005; Begg, Langley, & Williams, 1999; Koziol-McLain, Brand, Morgan, Leff, & Lowenstein, 2000; Norrish, North, Kirkman, & Jackson, 1994), including our previously published reliability study on the question in the written survey regarding motor vehicle crash injuries with telephone interviews (Alonso, Laguna, & Seguí-Gómez, 2006).

Validity of information sources regarding MVC is generally less than optimal. Ahlm, Eriksson, Lekander, and Björnstig analysed 580 officially registered traffic deaths in Sweden in 1999, they found that 490 were true accidental deaths, while 18 were suicides, 12 were deaths due to indeterminate causes, 59 were natural deaths and 1 case was not possible to evaluate due to missing data. Thus, only 84% of the officially registered 'accidental traffic deaths' were bona fide accidents (Ahlm et al., 2001). Landeira et al. compared the official coding of underlying cause of death on Death Certificates with coding based on primary data obtained from five hospitals and accident reports. Data showed an underreporting of deaths by traffic accidents, as 32% of the 50 deaths were coded as non-specific accidents (Ladeira & Gimarães, 1998). Grant,

*Corresponding author. Email: msegui@unav.es

ISSN 1745-7300 print/ISSN 1745-7319 online

DOI: 10.1080/17457301003728528
http://www.informaworld.com



International Journal of Injury Control and Safety Promotion
Publication details, including instructions for authors and subscription information:
http://www.informaworld.com/smpp/title~content=t713734346

Validation of self-reported motor-vehicle crash and related work leave in a multi-purpose prospective cohort

Juan Pons-Villanueva[a]; María Seguí-Gómez[b]
[a] Preventive Medicine and Public Health Department, School of Medicine, University of Navarra, Pamplona, Spain [b] Preventive Medicine and Public Health Department, European Center for Injury Prevention, School of Medicine, University of Navarra, Pamplona, Spain

First published on: 13 April 2010

To cite this Article Pons-Villanueva, Juan and Seguí-Gómez, María(2010) 'Validation of self-reported motor-vehicle crash and related work leave in a multi-purpose prospective cohort', International Journal of Injury Control and Safety Promotion,, First published on: 13 April 2010 (iFirst)
To link to this Article: DOI: 10.1080/17457301003728528
URL: http://dx.doi.org/10.1080/17457301003728528

Gregor, Beck, and Maio conducted a cross-sectional, observational study. Ninety-one patients transported by ambulance to an emergency department directly from the scene of a MVC were analysed. Agreement using κ between the crash investigation and different data sources was 0.58 for the police report, 0.33 for the ambulance report and 0.49 for the emergency department report (Grant et al., 2000). Similar problems exist in self-reported MVC. McGwin, Owsley, and Ball studied the agreement between self-reported MVC and state records, finding a κ index of 0.45 (McGwin, Owsley, & Ball, 1998). Anstety et al. found that the κ between retrospective self-reported crashes and retrospective state crashes was 0.18 and the κ for agreement between self-reported prospective crashes and prospective state crashes was 0.01 (Anstey, Wood, Caldwell, Kerr, & Lord, 2009).

Thus, the aim of this study was to continue our cohort data quality evaluation work, testing the reliability of the MVC-related questions in our cohort's questionnaire in a threefold manner: (a) reassessing the reliability of self-reported MVC with a larger sample size and a longer follow-up period; (b) assessing the reliability of self-reported work leave related to MVC; and (c) investigating the criterion validity of MVC and work leave by comparing the questions included in the questionnaire with the cohort participant's clinical notes (used as gold standard).

2. Methods

This study is based on the SUN cohort. This is an open-enrolment study of university graduates. A wide ranging questionnaire assessing exposures related to nutritional habits, physical exercise, road traffic and other factors is answered by participants of the study. Every 2 years participants are asked to answer short follow-up questionnaires asking about changes in exposure and health-related events (Seguí-Gómez et al., 2006) Participant's follow-up was done through mailed questionnaires. Information about previous MVC was gathered both from the baseline (Q_0) and the first two follow-up questionnaires (Q_2 and Q_4); this is, 2 and 4 years after the baseline questionnaire. The follow-up questionnaires included MVC-related questions, specifically in Q_2 two questions were posed: 'Since you answered the first questionnaire in this study, have you suffered any of these circumstances': (1) a MVC requiring hospitalisation of at least 24 h, (2) other MVC without hospitalisation? In Q_4 the questions were (1) the same as in Q_2 and (2) 'have you suffered other MVC without hospitalisation but with work leave?' Incidence of MVC and work leave was derived out of these questions. As of 3rd May 2005, 3507 participants had completed the second follow-up questionnaire (Q_4).

For this re-test reliability and criterion validity study we chose as inclusion criteria: (1) residents of Pamplona's metropolitan area, (2) participants who had stated not to have had any MVC hospitalisation in the baseline questionnaire (to avoid MVC's occurring prior to enrolment in the cohort), (3) participants who had not left blank the questions regarding MVC in either Q_2 or Q_4 follow-up questionnaires and (4) participants who had not left blank the question regarding MVC-related work leave in Q_4. There were 842 participants who fulfilled the selection criteria.

We invited the selected participants to participate in this reliability and validity study through letters (up to three consecutive letters were sent in case we had no answer to the first one). These letters were sent during 2005. In these letters, we asked them again about MVC and work leave to assess repeatability, specifically they had to check – if appropriate – the statement 'I confirm I had an MVC since I participated in the SUN study' and also they had to say for how long they had been on work leave. Patients were also asked for their consent to access their clinical notes, which we considered the gold standard, to validate their answers.

Surveillance of MVC can be done through different data sources. In any case it must be kept in mind that there is no perfect agreement between an event and the different information sources (self-reports, on-road assessment, ambulance records, emergency department records, hospital records, or state crash databases) and what actually happens. This is because some of the data sources are not necessarily involved in the process that evolves after an MVC and that some of them fail to register the mechanism of injury. National data sources provide the most thorough data for the general population. On the other hand, targeted cohorts on MVC allow for a deeper insight into events, risk factors and outcomes. There is a duty to study the validity of the source information of MVC events, as it can be concluded from the variability of agreement between different sources. Regarding self-reported MVC, the question is whether it really happened. The first way to assess this is to test–retest the participant asking him the same question twice. A second way is to assess whether the participant is really answering what is intended to be asked in the question (criterion validity) is contrasting the self-reported information to that of a gold standard. In this second aspect, there are several options which depend on what is taken as the gold standard. In this study, we took the clinical notes as our source of information.

We initially thought that the health centre's clinical notes would be the reference standard as Primary Health Care is a well established institution in Spain, a

country with a National Health Insurance system in place for several decades. The information system in most regions, including Navarra, where Pamplona is located, is integrated and electronic, making telemetric access available both to outpatient and in-hospital clinical notes and documents from 1999 onwards. In addition, sick (and maternity) leaves have to be signed by the primary health physician (or similar if the patient has work insurance). This is the case for both employees and self-employed workers. Therefore, it was expected that a patient's work leave would be recorded in the clinical notes, with the caveat that it should have been self-stated by the patient to the primary health care physician.

MVC, before a search through clinical notes, was defined in this study as any crash sustained as an occupant of a motor vehicle or as a pedestrian or cyclist being struck by a motor vehicle occurring during the same time frame in which the patient was participating in the cohort study. Before searching through clinical notes for evidence of MVC-related *work leave*, we defined it as an occupant of a motor vehicle or as a pedestrian or cyclist being struck by a motor vehicle and requiring at least a 1-day work leave permit.

Two trained research assistants working on a blind sample made a manual and systematic assessment of the clinical notes of those who gave consent.

Subjects lost to follow-up were excluded from the analysis. Differences between those who consented and those who did not were analysed with independent means' comparison and Pearson's χ^2 as convenient. We compared answers between written questionnaires and answers to this study's invitation letter, and between the former and the clinical notes in those who gave consent. Repeatability and criterion validity were assessed using Cohen's κ statistics. κ coefficients were labelled as suggested by Landis & Koch (1977). We calculated sensitivity, specificity and positive or negative values for the mailed questionnaire using the clinical notes as the gold standard. Results are graphically plotted as suggested by Bangdiwala, Haedo, Natal, & Villaveces (2008).

All statistical analyses were done using STATA/SE 9.

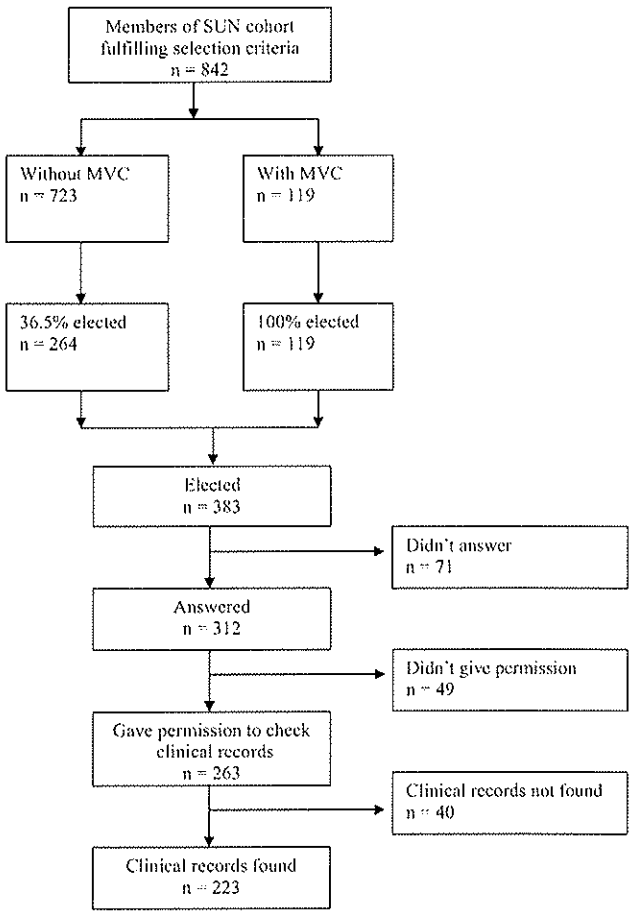
3. Results

As of June 3rd 2005, there were 4331 patients eligible for 4-year follow-up, from these 3507 had completed the second follow-up questionnaire (Q_4) (retention rate 80.9%) and 842 out of these fulfilled the selection criteria. From these, 119 had reported to have had an MVC after the beginning of the observation. Of the 723 that reported not having had an MVC a random sample of 264 was selected for this study (sampling rate 36.5%).

All of them were invited to enroll in the study (Figure 1). From these, 312 answered the letter (81.4% response rate) and 263 of these gave consent to access their clinical notes. Of those who consented, we were able to find clinical notes for 223 participants. Four participants returned their consent too late and are not included in our study (Figure 1).

Comparison between patients who answered the letter and those who did not shows that the proportion of patients who reported having had an MVC in the follow-up questionnaires was not different (Table 1).

Table 2 shows a similar comparison, but for those participants who answered the letter by stratifying whether they consented to the clinical note search. There were no differences in age or sex distribution. The proportion of patients who reported having had an MVC in the questionnaire was not different between those who gave consent and those who did not (Table 2).



SUN = Seguimiento Universidad de Navarra.
MVC = Motor vehicle crash

Figure 1. Study participation flow chart.

Re-test and criterion validity calculations were done weighting those who hadn't had an MVC or work leave as to represent the 100% of the original cohort's participants reporting not to have had this event. For re-test calculations the final weighted population was 696 for the MVC question and 620 for the work-leave question. For the criterion-validity analysis, the final weighted population was 497 for the MVC question and 446 for the work-leave question. Re-test reliability of the questionnaire with respect to the answer in the letter regarding MVC is summarised in Table 3. The percentage of agreement was 88.6%, with $\kappa = 0.55$. Sensitivity was 55.2%, specificity 95.3%. The positive predictive value was 70.3% and negative the predictive value was 91.4% (Table 3).

Table 4 summarises re-test reliability of the questionnaire with respect to the answer in the letter regarding work leave. The percentage of agreement was 97.1%, with $\kappa = 0.53$. Sensitivity was 45.8%, specificity 99.2%. Positive predictive value was 68.8% and negative predictive value was 97.8% (Table 4).

Criterion validity for MVC-related is detailed in Table 5. Percentage of agreement is 86.3% and Cohen's κ value is 0.25. Sensitivity was 45.7%, specificity 89.4%, positive predictive value was 24.6% and negative predictive value was 95.6% (Table 5).

Table 6 details the criterion validity for work leave. Percentage of agreement is 96.4% and Cohen's κ value is 0.25. Sensitivity was 37.5%, specificity 97.5%, positive predictive value was 21.4% and negative predictive value was 98.8% (Table 6).

These results are graphically represented in Figures 2 and 3. The agreement chart provides a visual assessment of agreement by comparing areas based on the cell frequencies from contingency tables. The row and column marginal totals determine rectangles within the larger square determined by the sample size. The larger the darkened area within the

Table 1. Sociodemographic characteristics of participants stratified between those participants who answered to the invitation letter and those who did not ($n = 383$).

	Did not answer, $n = 71$	Answered, $n = 312$	p (two-sided)
Age, mean (SD)	36.5 (1.09)	39.1 (0.5)	0.008
Gender			
Male	24	124	0.35
Female	47	188	
MVC			
No	43	221	0.09
Yes	28	91	

Table 2. Sociodemographic characteristics of participants stratified between those participants who consented to give access to their clinical notes and those who did not ($n = 312$).

	Did not consent, $n = 49$	Consented, $n = 263$	p (two-sided)
Age, mean (SD)	37.3 (9.6)	39.0 (9.7)	0.27
Gender			
Male	14	110	0.082
Female	35	153	
MVC			
No	38	183	0.26
Yes	11	80	

Table 3. Re-test assessment to the question in the invitation letter regarding MVC on all participants who answered to the invitation letter – weighted sample.

	Answer to the study invitation letter		
Answer to questionnaire	No	Yes	Total
No	553	52	605
Yes	27	64	91
Total	580	116	696

Percentage of agreement: 88.6%, κ: 0.55, Sensitivity: 55.2%, Specificity: 95.3%, Predictive value of positive: 70.3%, Predictive value of negative: 91.4%.

Table 4. Re-test assessment to the question in the invitation letter regarding MVC-related work leave on all participants who answered to the letter (within those reaching the 4-year follow-up) – weighted sample.

	Answer to the study invitation letter		
Answer to cohort questionnaire	No	Yes	Total
No	591	13	604
Yes	5	11	16
Total	596	24	620

Percentage of agreement: 97.1%, κ: 0.53, Sensitivity: 45.8%, Specificity: 99.2%, Predictive value of positive: 68.8%, Predictive value of negative: 97.8%.

Table 5. Criterion validity assessment of cohort questionnaire answers to MVC question in the follow-up questionnaires (Q_2 and Q_4) – weighted sample.

	Answer from clinical notes		
Answer to questionnaire	No	Yes	Total
No	413	19	432
Yes	49	16	65
Total	462	35	497

Percentage of agreement: 86.3%, κ: 0.25, Sensitivity: 45.7%, Specificity: 89.4%, Predictive value of positive: 24.6%, Predictive value of negative: 95.6%.

Table 6. Criterion validity assessment of questionnaire answers to MVC-related work leave question, within those who reached the 4-year follow-up – weighted sample.

	Answer from clinical notes		
Answer to questionnaire	No	Yes	Total
No	427	5	432
Yes	11	3	14
Total	438	8	446

Percentage of agreement: 96.4%, κ: 0.25, Sensitivity: 37.5%, Specificity: 97.5%, Predictive value of positive: 21.4%, Predictive value of negative: 98.8%.

rectangles, the larger the degree of agreement between the diagnostic test and the correct diagnosis.

4. Discussion

Given the administrative structure of medical electronic records, only participants living in Pamplona's metropolitan area were selected because direct access to their clinical notes was otherwise not possible. This selection could have been spread to other regions but we found this selection criteria provided enough subset sample for a re-test reliability and a criterion validity study.

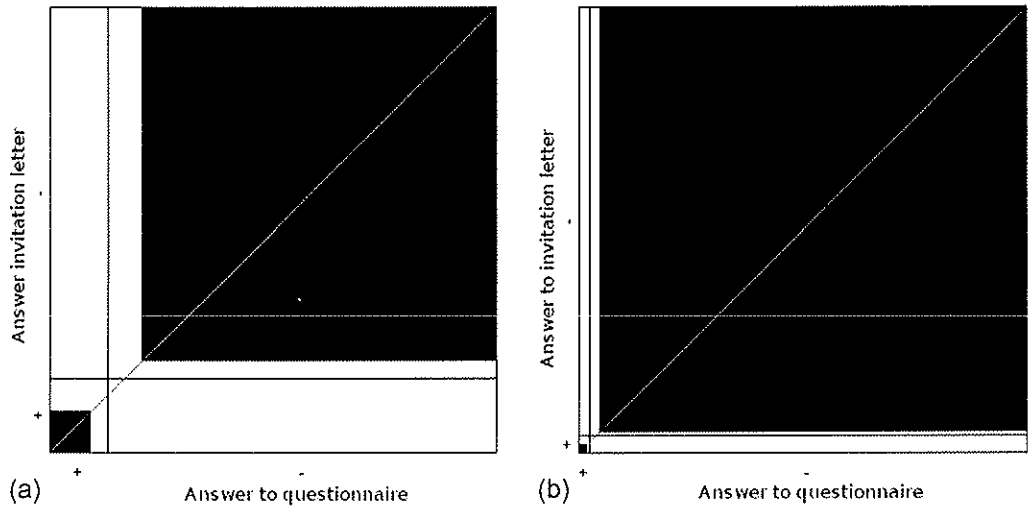


Figure 2. Test-retest study. (a) Agreement-diagnostic chart for the questionnaire answers to MVC (invitation letter answer as gold standard) (cf. Table 3). (b) Agreement-diagnostic chart for the questionnaire answers to MVC-related work leave (invitation letter answer as gold standard) (cf. Table 4).

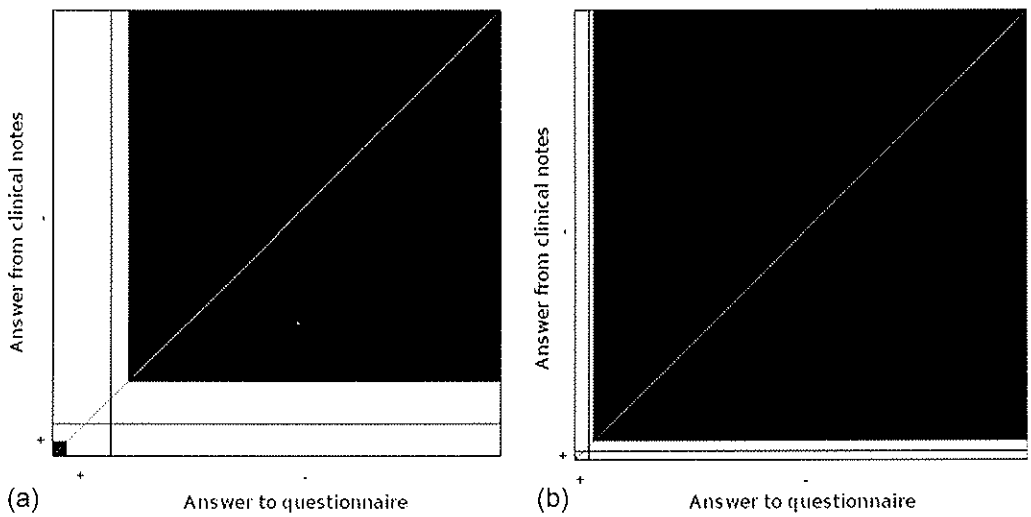


Figure 3. Criterion validation study. (a) Agreement-diagnostic chart for the questionnaire answers to MVC (clinical notes as gold standard) (cf. Table 5). (b) Agreement-diagnostic chart for the questionnaire answers to MVC-related work leave (clinical notes as gold standard) (cf. Table 6).

The proportion of reported MVC in the original cohort questionnaires was not statistically different between patients who answered the letter inviting them to join our data quality study and those who didn't answer (Table 1). Only one participant who had not answered the letter had abandoned the study.

The agreement charts are designed so that the frequencies in the diagonal cells from the contingency table determine darkened areas of perfect agreement within the rectangles and the unshaded areas within the rectangles represent the off-diagonal cells entries of disagreement (Bangdiwala et al., 2008). Thus, regarding test–retest calculations, there is more agreement in questions regarding work leave than in questions regarding MVC (Figure 2). The same happens in the criterion study (Figure 3).

However seemingly low the agreement values, these findings need to be put in perspective with others in the literature. Our repeatability findings on MVC are slightly lower, but consistent with those previously reported by us (Alonso, Laguna, & Seguí-Gómez, 2006) in a study where a subset of participants of the SUN cohort were assessed using a telephone interview. In this previously published study, we only included cases followed up for 2 years. The two questions assessed were 'Since you answered the first questionnaire in this study, have you suffered one of these circumstances: (1) a MVC injury requiring hospitalisation of at least 24 h (yes/no, if yes, tell us month and year); (2) other MVC injury not requiring hospitalisation (yes/no, if yes, tell us month and year)'. Our group found a Cohen's κ statistic of 0.63. Sensitivity was 83%, specificity 77%, positive predictive value 74% and negative predictive value 89%. Our current findings over a longer follow-up period yield slightly lower agreement values, possibly in relation to forgetfulness of the event with a longer follow-up period and the different source of information.

Interestingly, our current criterion validity findings are not that different from those published by others. Norrish et al. assessed validity of self-reported hospital admission with an admission computerised database. They found that only 58% of patients recalled all of their admissions and 16% recalled none of their admissions (Norrish et al., 1994). Arthur et al. presented data regarding self-reported MVC. κ index calculated from the data provided in the article is 0.25, a value not far from ours (Arthur et al., 2005). We found only two studies showing better values: Begg et al. found that 86% of the participants in their study were able to recall unintentional injury in the previous 3 years, having the health system database as their gold standard (Begg et al., 1999). Koziol-McLain et al. assessed re-test reliability of self-reported injury and found a Cohen's κ coefficient of 0.80 (95% CI 0.52–1.0) (Koziol-McLain et al., 2000).

Validation exercises in other exposures and events reveal interesting numbers and interpretations. For example, validation studies for other health topics within the SUN cohort present results similar to ours. Alonso et al. found an intraclass correlation coefficient of 0.35 when validating self-reported high-blood pressure (Alonso et al., 2000). Martinez-Gonzalez et al. found a κ index of 0.25 for self-reported physical activity during leisure time (Martínez-González et al., 2005). Only self-reported weight yielded very satisfactory results, with a κ index of 0.91 (Bes-Rastrollo, Pérez Valdivieso, Sánchez-Villegas, Alonso, & Martínez González, 2005).

Validation studies in the Nurses' Health Study for a dietary questionnaire yielded a mean of correlation coefficient between the dietary records and a questionnaire of 0.52 (Salvini et al., 1989). The same range of values is reported in the health professionals follow-up study, where they studied that correlations between two food-frequency questionnaires and diet records ranged from 0.45 to 0.74 (Hu et al., 1999).

Our current findings reveal that participants are not perfectly consistent in their reporting of (supposedly) major events, such as an MVC. Very interestingly, the self-report and official data sources do not agree perfectly either. This may be due to different factors worthy of debate.

The validity of the clinical notes should also be assessed. The fact that there is no perfect agreement between self-reported answers and reviewed clinical notes may be because health centre's clinical notes are indeed not a good gold standard. Interestingly, in Spain less than 50% of hospital admissions have E codes for mechanism of injury (Work Group on Traffic Accidents Impact Measure on Health in Spain, 2004), the WHO coding system meant to identify these events. Information on motor vehicle crash involvement is supposedly recorded in an administrative variable related to the source of payment (medical care for crash victims is meant to be covered by vehicle insurance instead of the national health system funds). Yet, this administrative variable is not included in the electronically available files that represent the medical history of a patient. Yet, hospitalisation data is, in principle, one of the very few data sources we can rely on to document this type of events. These facts emphasise the importance of registering hospital discharges as a means for further research.

Agreement between cohort participants and clinical notes is worse regarding work leave. Possible factors leading to this include that the primary health physician was not in charge of the participant's work leave paperwork or that self-employed participants

take 'official' work leave but actually go to work and this is what they report in their answers to us. Interestingly, many health interviews around the world use the work leave-related question to measure incidence of injuries in the population (Warner, Schenker, Heinen, & Fingerhut, 2005), although we are not aware of any specific validation exercise on these questions.

Another issue regarding validation of self-reported health events is the recall period. It has been previously thoroughly disclosed that the percentage of recall declines as time increases between event-time and recall-time (Harel et al., 1994; Warner et al., 2005). Yet, we would like to highlight that the longest recall period for all of our participants was 4 years.

It should be noted that the magnitude of κ is influenced by two factors. One is the prevelence index, which is the absolute difference proportion (between positive and negative) of agreeded classifications. If the prevalence index is high, chance agreement is also high, and κ is reduced accordingly. The effect of prevalence is greater for large values of κ than for small values. Second, the bias index, which is the difference proportion of disagreeded classifications. When there is a large bias index, κ is higher than when bias is low or absent. The effect of bias is greater when κ is small than when it is large (Byrt, Bishop, & Carlin, 1993; Feinstein & Cicchetti, 1990). In the present case, as seen in Table 7, there are both large prevalence indexes and low bias indexes, which should back a more optimistic interpretation of the κ values obtained. Furthermore, this conclusion is also suported by the difference between tha κ values and the maximum attainable κ values (Dunn, 1989).

In conclusion, even though our findings regarding κ and positive predictive values are below what one may have preferred, they seem to be within the bulk part of other validation and repeatability items often cited in the scientific literature. Thus, for the time being we propose to include this information in epidemiological studies to identify motor vehicle injury risk factors. Motor vehicle injuries amount to a large health burden to our society; pending future evidence that demonstrates a more efficient system to characterise and investigate them, we must rely on a combination of methods, including self-reported events, administrative databases and others.

Table 7. Agreement related measures for contingency Tables 3–6.

	κ	Prevalence index (%)	Bias index (%)	κ_{max}
Table 3	0.55	70.2	3.5	0.85
Table 4	0.53	93.5	1.2	0.79
Table 5	0.25	79.8	6.0	0.67
Table 6	0.25	95.0	1.3	0.72

κ_{max} = Maximum attainable κ value.

Acknowledgements

The authors thank Miguel Angel Martínez-González, principal investigator of the SUN study, Pedro Ruiz, María Alfonso and María Hidalgo for their clinical notes' data retrieval, Carmen de la Fuente for her work with questionnaires' management, Montse Ruiz for her manuscript review and Eduardo del Pozo de Dios for his graphical assistance. This study was funded by the Fondo de Investigación Santitaria (PI 03-030678).

References

Ahlm, K., Eriksson, A., Lekander, T., & Björnstig, U. (2001). All traffic related deaths are not "fatalities" – Analysis of the official Swedish statistics of traffic accident fatalities in 1999. *Lakartidningen, 98*, 2016–2022.

Alonso, A., Beunza, J.J., Delgado-Rodríguez, M., & Martínez-González, M.A. (2005). Validation of self-reported diagnosis of hypertension in a cohort of university graduates in Spain. *BMC Public Health, 5*, 94–100.

Alonso, A., Laguna, S., & Segui-Gómez, M. (2006). A comparison of information on motor vehicle crashes as reported by written or telephone interviews. *Injury Prevention, 12*, 117–120.

Anstey, K.J., Wood, J., Caldwell, H., Kerr, G., & Lord, S.R. (2009). Comparison of self-reported crashes, state crash records and an on-road driving assessment in a population-based sample of drivers aged 69–95 years. *Traffic Injury Prevention, 10*, 84–90.

Arthur, W. Jr., Bell, S.T., Edwards, B.D., Day, E.A., Tubre, T.C., et al. (2005). Convergence of self report and archival crash involvement data: A two-year longitudinal follow-up. *Human Factors, 47*, 303–313.

Bangdiwala, S.I., Haedo, A.S., Natal, M.L., & Villaveces, A. (2008). The agreement chart as an alternative to receiver-operating charasteristic curve for diagnostic tests. *Journal of Clinical Epidemiology, 61*, 866–874.

Barr, R.G., Herbstman, J., Speizer, F.E., & Camargo, C.A. (2002). Validation of self-reported chronic obstructive pulmonary disease in a cohort study of nurses. *American Journal of Epidemiology, 155*, 965–971.

Begg, D.J., Langley, J.D., & Williams, S.M. (1999). Validity of self-reported crashes and injuries in a longitudinal study of young adults. *Injury Prevention, 5*, 142–144.

Bes-Rastrollo, M., Pérez Valdivieso, J.R., Sánchez-Villegas, A., Alonso, A., & Martínez-González, M.A. (2005). Self-referred weight and body mass index validation in a Spanish graduates' cohort participants' (in spanish). *Revista Española de Obesidad, 3*, 352–358.

Byrt, T., Bishop, J., & Carlin, J.B. (1993). Bias, prevalence and kappa. *Journal of Clinical Epidemiology, 46*, 423–429.

Dunn, G. (1989). *Design and analysis of reliability studies: The statistical evaluation of measurement errors*. London, England: Edward Arnold.

Feinstein, A.R., & Cicchetti, D.V. (1990). High agreement but low kappa, I: The problems of two paradoxes. *Journal of Clinical Epidemiololgy, 43,* 543–549.

Grant, R.J., Gregor, M.A., Beck, P.W., & Maio, R.F. (2000). A comparison of data sources for motor vehicle crash characteristic accuracy. *Academic Emergency Medicine, 7,* 892–897.

Harel, Y., Overpeck, M.D., Jones, D.H., Scheidt, P.C., Bijur, P.E., Trumble, A.C., et al. (1994). The effects of recall on estimating annual nonfatal injury rates children and adolescents. *American Journal of Public Health, 84,* 599–605.

Hu, F.B., Rimm, E., Smith-Warner, S.A., Feskanich, D., Stampfer, M.J., Ascherio, A., et al. (1999). Reproducibility and validity of dietary patterns assessed with a food-frequency questionnaire. *American Journal of Epidemiology, 69,* 243–249.

Koziol-McLain, J., Brand, D., Morgan, D., Leff, M., & Lowenstein, S.R. (2000). Measuring injury risk factors: Question reliability in a statewide sample. *Injury Prevention, 6,* 148–150.

Ladeira, R.M., & Guimarães, M.D. (1998). Analysis of agreement between coding of underlying cause of death by traffic accidents. *Revista de Saúde Publica, 32,* 133–137.

Landis, J.R., & Koch, G.G. (1977). The measurement of observer agreement for categorical data. *Biometrics, 33,* 159–174.

Martínez-González, M.A., López-Fontana, C., Varo, J.J., Sánchez-Villegas, A., & Martínez, J.A. (2005). Validation of the Spanish version of the physical activity questionnaire used in the Nurses' Health Study and the Health Professionals' Follow-up Study. *Public Health Nutrition, 8,* 920–927.

McGwin, G. Jr., Owsley, C., & Ball, K. (1998). Identifying crash involvement among older drivers: Agreement between self-report and state records. *Accident; analysis and prevention, 30,* 781–791.

Norrish, A., North, D., Kirkman, P., & Jackson, R. (1994). Validity of self-reported hospital admission in a prospective study. *American Journal of Epidemiology, 140,* 938–942.

Salvini, S., Hunter, D.J., Sampson, L., Stampfer, M.J., Colditz, G.A., Rosner, B., et al. (1989). Food-based validation of a dietary questionnaire: The effects of week-to-week variation in food consumption. *International Journal of Epidemiology, 18,* 858–867.

Seguí-Gómez, M., de la Fuente, C., Vázquez, Z., de Irala, J., & Martínez-González, M.A. (2006). Cohort profile: The 'Seguimiento Universidad de Navarra' (SUN) study. *International Journal of Epidemiology, 35,* 1417–1422.

Stata/SE 9.0. *StataCorp LP.* 4905 Lakeway Drive. College Station, Tx 77845 USA.

Warner, M., Schenker, N., Heinen, M.A., & Fingerhut, L.A. (2005). The effects of recall on reporting injury and poisoning episodes in the National Health Interview Survey. *Injury Prevention, 11,* 282–287.

Willett, W., Stampfer, M., Chu, N.F., Spiegelman, D., Holmes, M., & Rimm, E. (2001). Assessment of questionnaire validity for measuring total fat intake using plasma lipid levels as criteria. *American Journal of Epidemiology, 154,* 1107–1112.

Work Group on Traffic Accidents Impact Measure on Health in SPAIN. (2004). *Study of mortality at 30 days by traffic accident (in Spanish).* Madrid: Ministry of Health.