IRTAD Conference 2026
Better Road Safety Data for Better Safety Performance
15-17 April 2026 Athens, Greece
International Transport Forum
National Technical University of Athens

What factors affect the match rates of police and emergency department data in Sweden's road traffic crash database Strada?

Christian Howard[1], Åsa Forsman

Traffic Safety and Traffic Systems, The Swedish National Road and Transport Research Institute (VTI), VTI SE-581 95 Linköping Sweden,
christian.howard@vti.se
asa.forsman@vti.se

Keywords: Road traffic safety; Police data; Hospital data; Data linkage; Strada; Match rate

Introduction

In road traffic safety, data linkage between police-based datasets and hospital-based datasets is valuable for understanding the contributing factors and injury outcomes of crashes. Police crash data often contain information on crash circumstances, but estimations of injury severity can be inaccurate since diagnosis and treatment of injuries is not the focus of the police. Conversely, hospital data can provide detailed injury information, but details on crash circumstances may be limited. Furthermore, the two sources may capture different parts of the road traffic crash population as neither source is complete. Several studies have shown the value of linked datasets to produce more complete and detailed datasets from which to develop interventions (Elvik and Mysen, 1999; Lombardi et al., 2022; Short and Caulfield, 2016; Soltani et al., 2022, Watson et al., 2015).

It is important to understand what factors influence successful record matches between a police and a hospital dataset. Incorrect matches (false positives) mean that injury outcomes are linked to the wrong crash circumstances. Conversely, matching failures (false negatives) mean that the counts are inflated as the same crash or injury will be counted more than once. Thus, much literature has been devoted to understanding the data linkage between police and hospital-based datasets (see Karimi et al., 2024, and Soltani, 2024, for recent overviews).

Strada (the Swedish national registry of road traffic crashes) has been mostly absent from the literature on data linkage. One feature of Strada in that both the police and all hospital emergency departments (EDs) in Sweden report directly to the system. This means that the EDs not only provide AIS-coding of injuries but also information on crash circumstances (such as location, time, crash type and the patient's own account of what happened). Reports are matched by the system continuously and the resulting linked dataset is made available to selected government agencies through an online portal.

The primary aim of this study is to investigate what factors affect the match rates between police and emergency department reports in Strada. Furthermore, a secondary aim is to provide information and insights regarding a national data linkage system that is essentially real-time.

Methods

The study is based on police and ED data on injured persons reported to Strada in 2022 and 2023. Since June 2021, ED reporting to Strada has been mandatory by law. The number of EDs in Sweden has fluctuated over time but is today close to 70. The EDs are obliged to report all persons injured in a road traffic crash. The police should, according to law, report all crashes that come to their attention and that involve at least one injured road user. The final data set, after applying some exclusion criteria, includes 18,882 records only reported by the police, 40,500 records only reported by EDs, and 9,954 records reported by both sources.

The Strada data linkage method uses personal identity numbers and time of crash to match police records with ED records. If the personal identity numbers are incomplete or missing, the records can still be identified as belonging to the same crash, depending on the distance and time, but not identified as being the same person.

[1] * Corresponding author. Tel.: +46-768-584-605;
E-mail address: christian.howard@vti.se

In this study, only linking of individuals are considered. Linkage performance is measured as the *police match rate* (the number of linked records divided by the total number of police records) and the *ED match rate* (the number of linked records divided by the total number of ED records).

A logistic regression model is applied to investigate what factors influence the match rate. The factors considered are *injury severity*, *sex*, *age group*, *road user category*, and *county*.

Results and conclusions

The results show that the police match rate is estimated to be 0.345 and the ED match is estimated to be 0.197. This difference can most likely be attributed to differences in police and ED coverage, where police underreporting of cyclist crashes is well known from the literature and cyclists represent more than half of the ED reported persons in Strada. The logistic regression results show that ED reported cyclists had the lowest likelihood of being matched with a police record out of all ED reported road users. Regarding the police match rate, motorcycle riders stand out as having the highest likelihood of a record linked to an ED record. Bus occupants have both a low police match rate and a low ED match rate, indicating that neither source may provide sufficient coverage of these road users.

The results further show that higher injury severities are associated with higher match rates. This is consistent with the previous literature (Alsop and Langley, 2001; Amoros et al., 2006; Rosman, 2001; Short and Caulfield, 2016; Soltani et al. 2022; Wilson et al., 2012) and is most likely because higher severity crashes have a higher degree of both police and ED involvement. The results also show that both the police match rate and ED match rate can vary significantly between counties. The possible presence of regional differences should be reflected on when studying the data linkage literature as many previous studies focus on one particular region or hospital.

References

Alsop, J., Langley, J. (2001). Under-reporting of motor vehicle traffic crash victims in New Zealand. Accident Analysis & Prevention, 33 (3), 353-359. https://dx.doi.org/10.1016/S0001-4575(00)00049-X.

Amoros, E., Martin, J.-L., Laumon, B. (2006). Under-reporting of road crash casualties in France. Accident Analysis & Prevention, 38 (4), 627-635. https://dx.doi.org/10.1016/j.aap.2005.11.006.

Elvik, R., Mysen, A. (1999). Incomplete Accident Reporting: Meta-Analysis of Studies Made in 13 Countries. Transportation Research Record: Journal of the Transportation Research Board, 1665 (1), 133-140. https://dx.doi.org/10.3141/1665-18.

Karimi, S., Hosseinzadeh, A., Kluger, R., Wang, T., Souleyrette, R., Harding, E. (2024). A systematic review and meta-analysis of data linkage between motor vehicle crash and hospital-based datasets. Accident Analysis & Prevention, 197, 107461. https://www.sciencedirect.com/science/article/pii/S000145752400006X.

Lombardi, L. R., Pfeiffer, M. R., Metzger, K. B., Myers, R. K., Curry, A. E. (2022). Improving identification of crash injuries: Statewide integration of hospital discharge and crash report data. Traffic Injury Prevention, 23 (sup1), S130-S136. https://dx.doi.org/10.1080/15389588.2022.2083612.

Rosman, D. L. (2001). The Western Australian Road Injury Database (1987–1996). Accident Analysis & Prevention, 33 (1), 81-88. https://dx.doi.org/10.1016/S0001-4575(00)00018-X.

Short, J., Caulfield, B. (2016). Record linkage for road traffic injuries in Ireland using police hospital and injury claims data. Journal of Safety Research, 58, 1-14. https://dx.doi.org/10.1016/j.jsr.2016.05.002.

Soltani, A., Edward Harrison, J., Ryder, C., Flavel, J., Watson, A. (2024). Police and hospital data linkage for traffic injury surveillance: A systematic review. Accident Analysis & Prevention, 197, 107426. https://dx.doi.org/10.1016/j.aap.2023.107426.

Soltani, S., Schwarcz, L., Morris, D., Plevin, R., Dicker, R., Juillard, C., Nwabuo, A., Wier, M. (2022). What is counted counts: An innovative linkage of police, hospital, and spatial data for transportation injury prevention. Journal of Safety Research, 83, 35-44. https://dx.doi.org/10.1016/j.jsr.2022.08.002.

Watson, A., Watson, B., Vallmuur, K. (2015). Estimating under-reporting of road crash injuries to police using multiple linked data collections. Accident Analysis & Prevention, 83, 18-25. https://dx.doi.org/10.1016/j.aap.2015.06.011.

Wilson, S. J., Begg, D. J., Samaranayaka, A. (2012). Validity of using linked hospital and police traffic crash records to analyse motorcycle injury crash characteristics. Accident Analysis & Prevention, 49, 30-35. https://dx.doi.org/10.1016/j.aap.2011.03.007.