WARSAW UNIVERSITY OF TECHNOLOGY Index 351733
FACULTY OF CIVIL ENGINEERING
COMMITTEE FOR CIVIL AND WATER ENGINEERING
POLISH ACADEMY OF SCIENCES ISSN 1230-2945

DOI: 10.24425/ace.2021.138051
ARCHIVES OF CIVIL ENGINEERING
Vol. LXVII ISSUE 3 2021

s. 197–211

Research paper

Identifying factors and conditions contributing to cyclists' serious accidents with the use of association analysis

H. Anysz[1], P. Włodarek[2], P. Olszewski[3], S. Cafiso[4]

Abstract: Being negatively impressed by the data published by the European Commission in CARE (Community database on Accidents on the Roads in Europe), where Poland is presented as the European Country with the highest rate of fatalities in road crashes involving cyclists during 4 years period (2009–2013), the Authors decided to analyse available data. Bikes become a more and more popular means of transport and the way of active recreation. In Warsaw, the share of bicycle trips rises 1 to 3% per year. The aforementioned, together with increasing traffic density, caused 4233 registered injuries among cyclists in 2018 in Poland. In 286 cases the accidents were direct reasons for the cyclists' death. Considering these facts, it becomes extremely important to point the most influencing factors and conditions contributing to cyclists' serious accidents. One-dimensional or two-dimensional statistics are not sufficient to find all important associations between the road conditions and the number of cyclists' accidents. To overcome that the association analysis is applied. The results of the analysis can contribute to increasing the knowledge and safety of transport.

Keywords: cyclist accidents, bicycle traffic, cyclist injuries, association analysis, market basket analysis

---

[1] PhD., Eng., Warsaw University of Technology, Faculty of Civil Engineering, Al. Armii Ludowej 16, 00-637 Warsaw, Poland, e-mail: h.anysz@il.pw.edu.pl, ORCID: https://orcid.org/0000-0002-3804-5859

[2] MSc., Eng., Warsaw University of Technology, Faculty of Civil Engineering, Al. Armii Ludowej 16, 00-637 Warsaw, Poland, e-mail: p.wlodarek@il.pw.edu.pl, ORCID: https://orcid.org/0000-0003-2680-3138

[3] DSc. PhD., Eng., Warsaw University of Technology, Faculty of Civil Engineering, Al. Armii Ludowej 16, 00-637 Warsaw, Poland, e-mail: p.olszewski@il.pw.edu.pl, ORCID: https://orcid.org/0000-0003-3470-4471

[4] Prof., University of Catania, Department of Civil Engineering and Architecture, Viale Andrea Doria 6, 95131 Catania, Italy, e-mail: dcafiso@unict.it, ORCID: https://orcid.org/0000-0002-7247-0365

1. Introduction

1.1. Bicycle traffic and accidents

In Poland, bicycle traffic is on the rise. A few millions of Poles cycle – not just to work or shop, but also to get around and as a form of pastime. People are starting to enjoy cycling – over the last few years the share of bicycle trips in Warsaw has almost doubled (from 3,8% in 2015 to 7,5% in 2019 [1]). There is a lot of improvements that can increase cycling and one of the most important [2] could be the construction of infrastructure for bicycles. During the last nine years, the length of paths for bicycles in Warsaw has doubled (Fig. 1).

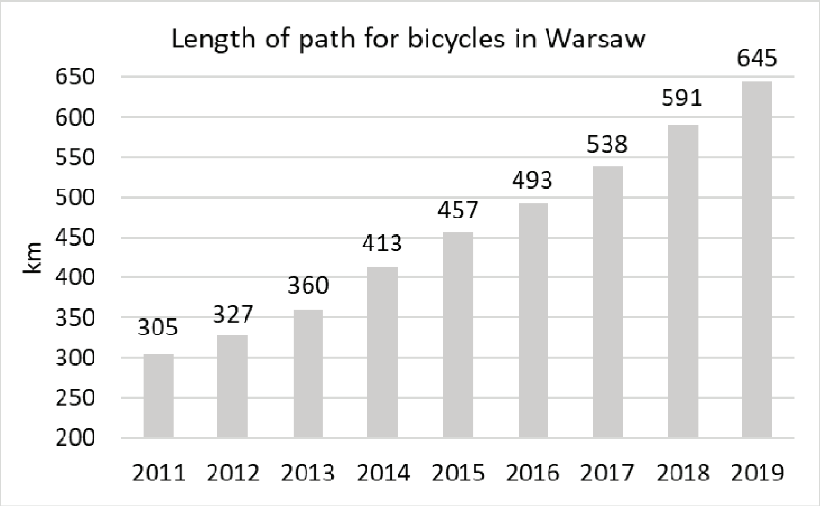
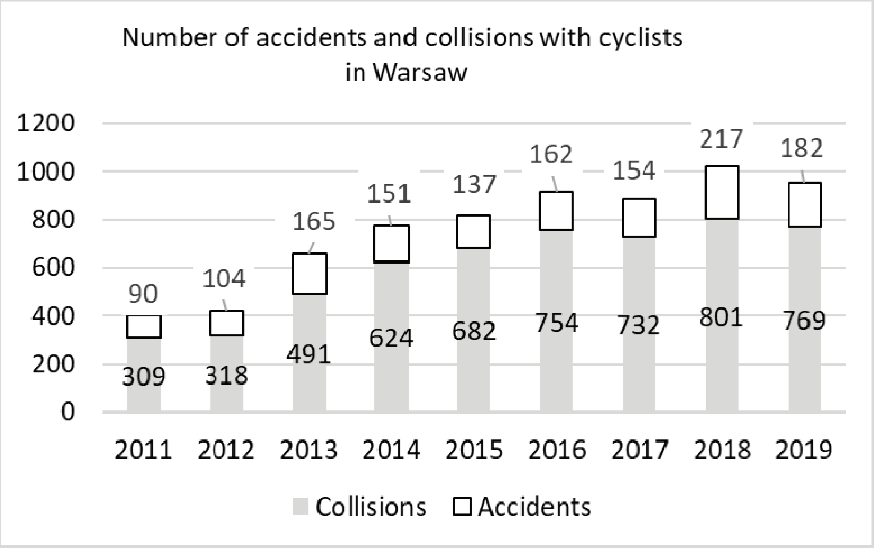


Fig. 1. Length of paths for bicycles in Warsaw [1]

Unfortunately, there is also an increase in the number of accidents and collisions involving cyclists. Between 2011 and 2019 there were twice as many accidents and over two times as many collisions. A similar increase can also be observed among the number of injured cyclists in Warsaw (between 2011 and 2019 there were twice as many slight and serious injuries) and we have between 1 and 5 fatalities every year. If we assume that bicycle traffic volume is approximately proportional to the length of bicycle paths in Warsaw, then increase of accidents can be explained by the growth in bicycle traffic – both numbers increase about 2 times in 8 years. However, any increase in the number of accidents is worrying regardless of the underlying reasons and should be investigated with the aim of identifying the conditions which could be changed to improve safety.



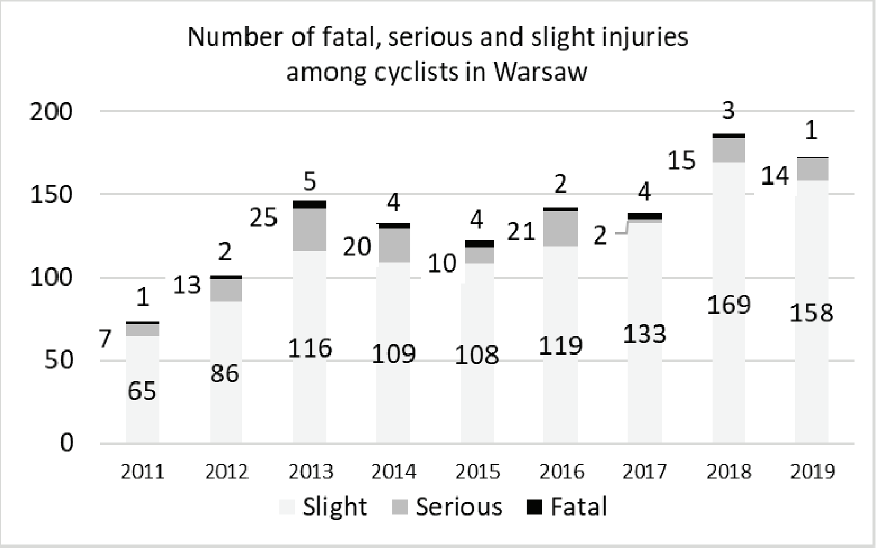


Fig. 2. Number of accidents and collisions with cyclists in Warsaw (higher) and fatal, serious or slight injuries of cyclists in Warsaw (lower) [3]

In Poland, cycling in cities takes place mainly on separate cycle tracks, outside cities it is most often shared with car traffic, less often on separate cycle tracks. Accident statistics for the whole country are also very bad, the average number of cyclist fatalities in Poland in 2011–2019 was about 280, with about 4,100 injuries. The downward trend is very slight.

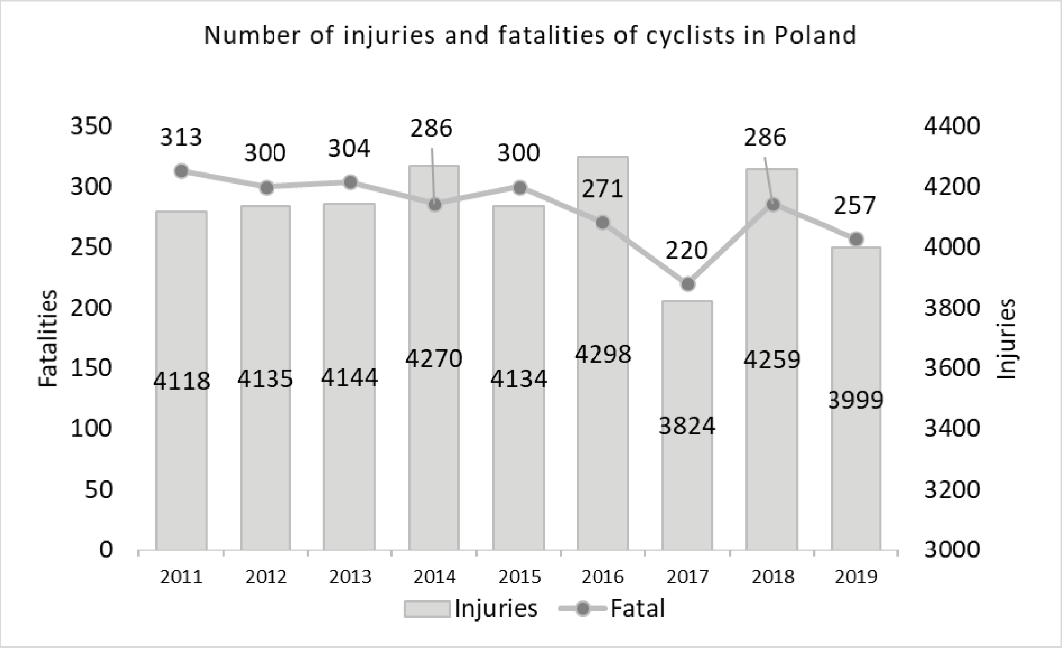


Fig. 3. Number of fatalities and injured cyclists in Poland [3]

Accidents involving cyclists are very scattered, thus standard epidemiological analyses sometimes are very difficult and could not give satisfactory results. Some researchers use conflict technique to find causes of accidents or their circumstances but it could be used only locally or on specific parts of the infrastructure (e.g. on cycle track crossings [4]).

1.2. Association analysis

Association analysis is also called market basket analysis, as it was invented to enhance sales of supermarkets through analysis of the content of shoppers' trolleys. Applying this tool, consumers' habits can be found [5]. The rules extracted from shoppers' receipts are analysed, and based on them, the locations of shelves in a supermarket can be adjusted or the pricing strategy of goods offered there. The idea of market basked analysis is to find the frequent, simultaneous appearance of goods or phenomena (e.g. brown shoes and brown socks or bread and butter) [6, 7]. There are numerous examples of applications of this type of analysis e.g. in biology [8], in education [9]. If time is introduced to this analysis, then there are many more possible applications of association analysis to utilise. Applications are obvious in medicine, where some jointly appearing phenomena (predecessors), called there symptoms, are causing the effect (consequent) – disease, illness [10]. Also, predictions can be done based on sequential association analysis [11]. Finding the rules between predecessors and consequents enabled the application of market basket analysis at areas closely related to civil engineering e.g. in quality management [12], collusion in road construction tenders [13], client's risk assessment in the road construction sector [14]. The feature of the rule finding of market basket analysis was recognized also by road safety researchers [15, 16, 17]. Market basket analysis empowers finding the rules, associations between predecessors (processes, facts, states, phenomena) that influence the analysed output – called a consequent [6, 7]. It helps to answer the question which combinations of predecessors (body b) make the consequent (head h) more likely to happen. The rules, expressed as "if body than head" ($b \longrightarrow h$) usually have different strength. It is described by two main ratios called support (sup) and confidence (conf) defined as:

$$\sup(b \rightarrow h) = \frac{N(b \cap h)}{n} \tag{1.1}$$

$$conf(b \rightarrow h) = \frac{N(b \cap h)}{N(b)} \tag{1.2}$$

where:
n – number of all analysed cases, $N(b \cap h)$ – number of cases where the body and the head appear simultaneously, $N(b)$ – number of cases where the body appears

The supplementary ratio – called "lift" – is often used when sequential association analysis is applied. It is defined as:

$$lift(b \rightarrow h) = \frac{conf(b \rightarrow h)}{P(h)} \tag{1.3}$$

where:
$P(h)$ – probability of the head appearance

This probability can be calculated as the ratio of the number of cases where the head appears to the total number of cases. Considering the rules for which the lift is greater than one protects from the situation where the rule of high confidence e.g. 80 % is analysed, but the probability of the head appearance (independent from the body appearance) is 90 %. For the described case the lift value is below 1, so the rules are only meaningful where the lift is greater than 1. The lift, in fact, depends also on the size of the database used for the analysis. It is obvious that – independently from the lift value – the rules with the confidence lower than 50 % are meaningless. Table 1 is presented for easier understanding of the meaning of lift. There are a few examples of values of sup, conf, and lift for some rules found in the created databases of different sizes.

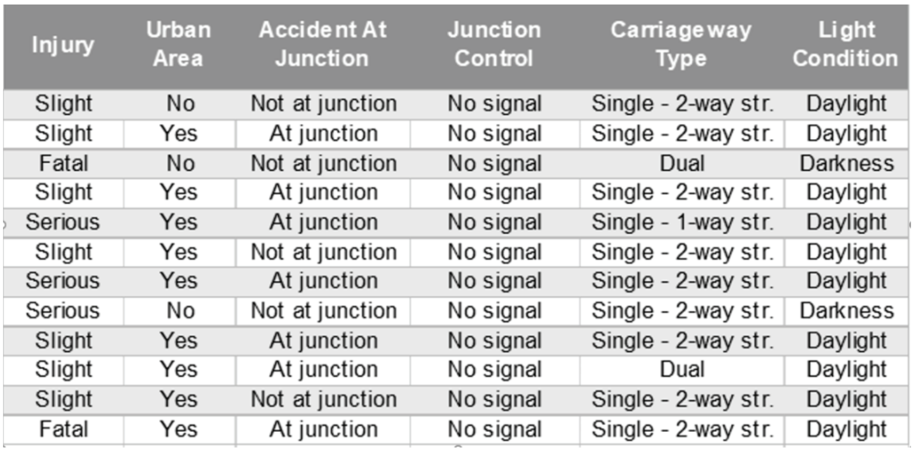
Table 1. Examples of calculations of support, confidence, and lift (for the sample sets of data)

Total number of cases	Number of cases with *b*	Number of cases with *h*	Number of cases where *b*→*h*	*P*(*h*)	sup(*b*→*h*)	conf(*b*→*h*)	lift(*b*→*h*)
100	40	40	20	0,4	20 %	50 %	1,250
100	40	40	30	0,4	30 %	75 %	1,875
1000	400	30	30	0,03	3 %	7,5 %	2,5
1000	40	30	30	0,03	3 %	75 %	25
1000	80	40	30	0,03	3 %	37,5 %	12,5

The examples presented above prove that the meaningful rule should have conf above 50 % and lift above 1. These simple calculations do not require computer support (for a reasonably low number of cases). However even then, if the body consists of several features, as well as the head, the task of exploring the database to find the meaningful rules becomes very time-consuming even for an advanced personal computer. Let's assume that there are 10 features describing a certain phenomenon. There are two decisions to undertake: how many features to consider, and which of them to consider in the analysis. There are 1013 possibilities. Moreover, considering that market basket analysis works better based on binary data, the thresholds for each feature separately should be set. If only 10 possible values of the thresholds are summed for one feature, when 10 features are considered, it gives 1E10 possible bodies to analyse. For the analysis presented in the following sections Statistica 13.1 software (by Dell) with the built-in association analysis module is applied. These kind of tools have been used in the area of road safety analysis [15, 16, 17]. Other tools also are applied to assess road traffic safety e.g. multi-assessment decision modelling [18] or logistic regression [19]. Cited studies concentrate on car as the main traffic mode. The present analysis is focused on cyclists' safety.

2. Database

Based on information provided by the Polish Motor Transport Institute that can be found at [20], a database concerning injuries of cyclists involved in road accidents was created. It comprises the following types of data: cyclist's injury severity, general location (an urbanised area or not), specific location (at a junction or not), the existence of traffic signals at intersections, light condition (day, night, twilight) and type of road (single or dual carriageway). There were 21.470 accidents with cyclists recorded in Poland in the period from 2009 to 2013. The original database (an example is presented in fig. 4) is transformed into the form suitable for market basket analysis with only binary data, coded as False=0 and True=1. That form is presented in tab. 2.



Injury	Urban Area	Accident At Junction	Junction Control	Carriageway Type	Light Condition
Slight	No	Not at junction	No signal	Single - 2-way str.	Daylight
Slight	Yes	At junction	No signal	Single - 2-way str.	Daylight
Fatal	No	Not at junction	No signal	Dual	Darkness
Slight	Yes	At junction	No signal	Single - 2-way str.	Daylight
Serious	Yes	At junction	No signal	Single - 1-way str.	Daylight
Slight	Yes	Not at junction	No signal	Single - 2-way str.	Daylight
Serious	Yes	At junction	No signal	Single - 2-way str.	Daylight
Serious	No	Not at junction	No signal	Single - 2-way str.	Darkness
Slight	Yes	At junction	No signal	Single - 2-way str.	Daylight
Slight	Yes	At junction	No signal	Dual	Daylight
Slight	Yes	Not at junction	No signal	Single - 2-way str.	Daylight
Fatal	Yes	At junction	No signal	Single - 2-way str.	Daylight

Fig. 4. Example of the original database

The database contains redundant information in the case of three descriptors of the single phenomenon (e.g. light condition: daylight, twilight, darkness). It would be sufficient to provide the value of two of them. The third descriptor is remained to simplify the form of the rules found in case of negation (e.g. daylight=0 and twilight=0 means darkness=1).

Table 2. The format of the transformed database

Bodies									Heads		
Urban area	Signalling	Junction	Single 2-way	Single 1-way	Dual roadway	Daylight	Twilight	Darkness	Fatal injuries	Serious injuries	Slight injuries
0	0	0	1	0	0	1	0	0	0	0	1
1	0	1	1	0	0	1	0	0	0	0	1
0	0	0	0	0	1	0	0	1	1	0	0
1	0	1	1	0	0	1	0	0	0	0	1
1	0	1	0	1	0	1	0	0	0	1	0
1	0	0	1	0	0	1	0	0	0	0	1
1	0	1	1	0	0	1	0	0	0	1	0
0	0	0	1	0	0	0	0	1	0	1	0
1	0	1	1	0	0	1	0	0	0	0	1
1	0	1	0	0	1	1	0	0	0	0	1
1	0	0	1	0	0	1	0	0	0	0	1
1	0	1	1	0	0	1	0	0	1	0	0

3. Database exploration

3.1. Basic analysis

There are 5 features describing the conditions associated with each accident. They cannot be presented at one time. Table 3 contains the data concerning accidents which happened at junctions.

Table 3. The share of injuries by level for different conditions, occurring at junctions

Conditions of accidents	Share of injuries in %			Number of accidents
	Fatal	Serious	Slight	
Urban area	3,9	23,9	72,2	7428
Non-urban area	19,3	29,2	51,5	487
Junction without signals	5,2	24,4	70,3	6652
Junction with signals	2,9	23,1	74,0	1263
Single 2-way	5,3	24,8	69,9	6273
Single 1-way	2,4	19,6	78,0	372
Dual roadway	3,5	22,6	73,9	1270
Daylight	4,3	24,2	71,4	6827
Twilight	7,6	26,5	65,9	393
Darkness	8,3	23,0	68,6	695

There were 1571 accidents with fatal consequences, but only 24,5% of them were located at junctions (table 4). From the total number of accidents 21470, 63% (i.e. 13555) took place not at junctions. The shares of different levels of injuries for accidents located not at junctions are presented in table 5.

Table 4. The shares and numbers of accidents with fatal injuries

The share of accidents	At junction	Not at junction
Urban area	18,5 % 290 cases	36,7 % 577 cases
Non-urban area	6 % 94 cases	38,8 % 610 cases
Total	24,5 %	75,5 %

Table 5. The shares of injuries by level for different conditions, occurring not at junctions

Conditions of accidents	Share of injuries in %			Number of accidents
	Fatal	Serious	Slight	
Urban area	5,5	26,9	67,5	10416
Non-urban area	19,4	28,1	52,4	3139
Single 2-way	9,3	27,1	63,6	12252
Single 1-way	2,5	21,0	76,5	396
Dual roadway	4,1	30,9	65,0	907
Daylight	6,5	26,8	66,8	10788
Twilight	15,2	27,0	57,7	880
Darkness	18,8	29,7	51,5	1887

The shares of the level of severity differ between urban and not-urban area as presented in figure 5.

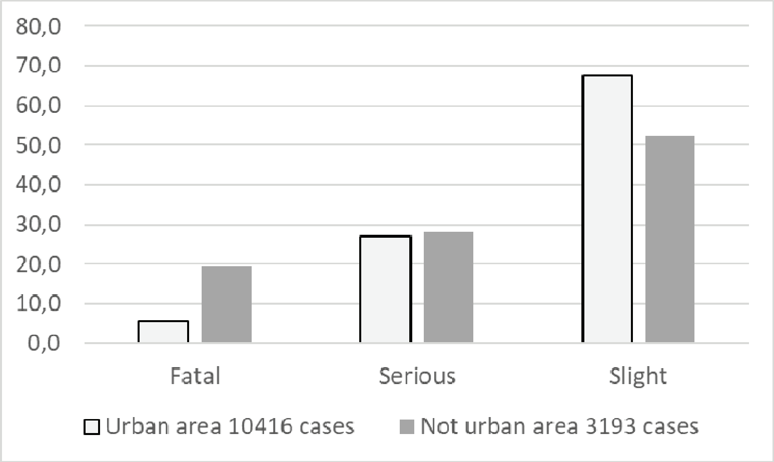


Fig. 5. The shares of injuries by level for urban and non-urban areas, at junctions
(each type of area sums up to 100 %)

To increase the readability of the factors' distribution, 3-D plot can be drawn for selected features. The light condition and place of crash are combined in figure 6 for fatal injuries.

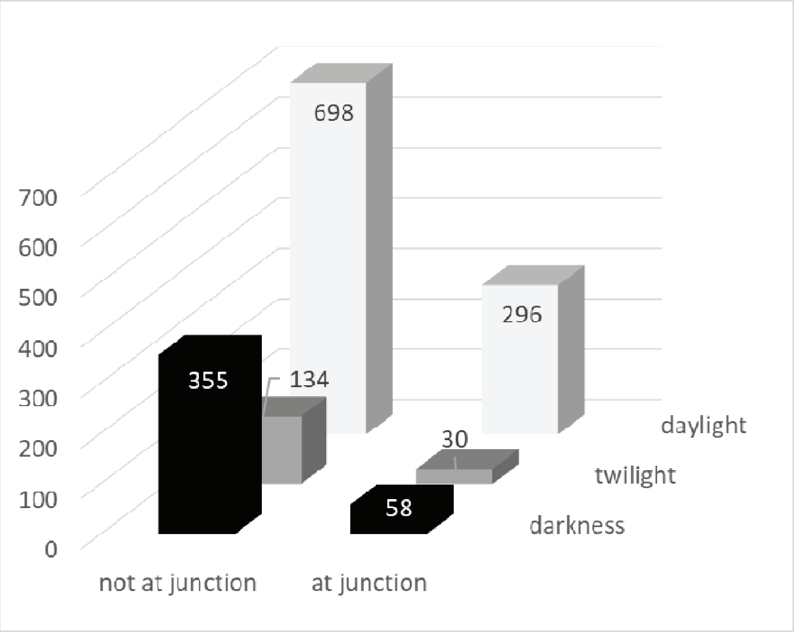


Fig. 6. Number of fatal injuries, at junctions and not at junctions

As presented in figure 5, the share of fatal and serious injuries is bigger outside urban areas. Based on that, it was decided to search for the rules splitting the dataset in two parts: urban and non-urban crashes.

3.2. Finding the rules with market basket analysis

To find the meaningful rules, the following preconditions were defined (i.e. set in the searching rules of the software applied): confidence over 50% and support over 0,2%. The rules with confidence lower than 50 % do not bring any explanation to the phenomenon examined (except information that the head of rule depends on other bodies than the one chosen for the analysis). It may seem that support equal to 0,2% is very low, but given the size of the database i.e. 21470, it amounts to almost 43 cases. When combined with high confidence, it can be a very meaningful rule. All the rules presented below have the lift over 1.

Searching the database (described in section 3.1) for the rules concerning fatal injuries has produced a low number of meaningless rules. Therefore, it was decided to combine the categories “serious” and “fatal” injuries into one class, marked as FS (fatal or serious injuries). Despite that, no rules for FS are found in urban areas. However, for the other class (slight injuries, FS=0) rules are detected and presented in table 6.

Table 6. The rules found for slight injuries, for accidents in an urban area

Rule number	Body (if)	Head (then)	conf %	sup %
1	junct=1	FS=0	72,3	30,0
2	junct=1 and signal=1	FS=0	74,2	5,2
3	junct=1 and signal=1 and dual=1	FS=0	75,0	2,7
4	junct=1 and signal=1 and dual=1 and day=1	FS=0	75,8	2,3

Rule number 1 can be read: if the accident is at a junction (in an urban area), there is a 72,3% chance, that cyclist's injuries are only slight. The support for this rule is 30,0%. It means that the rule is confirmed by 5353 cases (30% out of 17844 which happened in an urban area). It is a very strong rule. Its general validity is confirmed also by the fact that giving the conditions of an accident more specific (as in rules 2, 3, and 4), the confidence of the rule rises and the support of the rule lowers. The rules for non-urban area and slight injuries are also identified. They are presented in table 7.

Table 7. The rules found for slight injuries, for accidents in non-urban area

Rule number	Body (if)	Head (then)	conf %	sup %
5	junct=1 and signal=1	FS=0	65,2	41,4
6	junct=1 and signal=1 and dual=0	FS=0	75,0	33,1
7	junct=1 and signal=1 and dual=0 and day=1	FS=0	66,7	22,1

Comparing rule 5 (concerning non-urban area) and the rule 1 (concerning an urban area), it can be stated that the road traffic participants (drivers, bikers, pedestrians) easily recognize junctions in an urban area and other factors than listed in the database (e.g. drivers concentration, speed of vehicles) make cyclists' injuries slight when they happen there. Junctions outside urbanized area are recognized (as town junctions) only if they are equipped with traffic signals. Then rule 5 is similar to rule 1. It tells that if junctions outside urbanized area is equipped with traffic signals, slight injuries of a cyclist can be expected in 65,2% of accidents. The rule is strong, as it is supported by 1501 cases (41,4% out of 3626 cases in non-urban area).

The statement about low importance of traffic signals in an urban area during the day can be proved by comparing the following rules, presented in table 8.

Table 8. The rules found for slight injuries for accidents in an urban area in daylight

Rule number	Body (if)	Head (then)	conf %	sup %
8	junct=1 and signal=0 and day=1	FS=0	72,1	21,7
9	junct=1 and signal=1 and day=1	FS=0	74,4	4,4

Rules 8 and 9 differ only with the existence of traffic signal, but the confidence of appearance of only slight injuries of a cyclist in an accident is very close.

There are also 3 strong rules found for FS (serious or fatal) injuries outside an urban area. They are presented in table 9.

Table 9. The rules found for FS injuries for accidents in a non-urban area.

Rule number	Body (if)	Head (then)	conf %	sup %
10	junct=1 and dual=1 and day=0	FS=1	80,0	0,2
11	junct=0 and dual=1 and day=0	FS=1	64,2	0,4
12	dual=1 and day=0	FS=1	70,1	0,2

The rules presented above suggest that there are two key parameters (lack of daylight and dual roadway), if they appear simultaneously outside an urban area, it makes the risk of FS injuries of a cyclist in an accident high. These two parameters appear jointly in all three rules found for non-urban area found for FS=1 (i.e. in rules number 10, 11, 12). It can be stated – based on rule 12 – that if it is not a day, and the accident takes place outside an urban area on a dual roadway, the risk of fatal or serious injuries of a cyclist is 70%. The risk is even higher (80%) if it happens at a junction. If the accident is not at a junction the risk lowers a bit, to 64,2 %. The importance of daylight is emphasised by the rules for slight injuries (rules number 8 and 9). They are supported by the body day=1.

4. Conclusions

The shares of fatal and serious injuries in urban and non-urban areas differ and therefore separate analyses are suggested. Only 13% of fatal injuries located in non-urban areas were at junctions (94 out of 704), while in urban areas 33% (290 out of 867) were at junctions. No rules with confidence higher than 50% were found for fatal and serious injuries accidents in urban areas. An unexpected result is that, the daylight makes the existence of traffic signals practically not influencing the share of slight injuries in urban area. On the other hand, there are no meaningful rules for FS injuries in an urban area. For non-urban area, the key parameters (if appearing jointly) make the risk of FS injuries high i.e. lack of daylight and dual roadway. This kind of analysis, taking into consideration two or more features influencing a given type of injury is possible with a spreadsheet, but it is significantly time-consuming then. Without software supported association analysis, it is very difficult to find how many features are important and which of them should be considered. Association analysis allows automatically finding which set of the features supports the rules. The proposed method is especially efficient when applied to databases with multiple features. Strong rules (with high confidence) are found together with their importance. Comparing the rules (even the weak ones) can help to understand which combination of factors and conditions of the road system should be changed to reduce the number and severity of crashes involving cyclists. Including in the database other features (e.g. type of infrastructure for cyclists, night street lightning, personal lights of cyclists, weather conditions) could give more rules and increase our knowledge of the cyclist safety phenomenon.

References

[1] Warsaw Cycle Report website: http://transport.um.warszawa.pl/ruch-rowerowy/raporty-rowerowe

[2] N. Stamatiadis, S. Cafiso and G. Pappalardo, A Comparison of Bicyclist Attitudes in Two Urban Areas in USA and Italy, The 4th Conference on Sustainable Urban Mobility, pp. 272–279, 2018. https://doi.org/10.1007/978-3-030-02305-8_33

[3] Police website: http://statystyka.policja.pl/st/ruch-drogowy/76562,wypadki-drogowe-raporty-roczne.html

[4] P. Włodarek, P. Olszewski, Traffic safety on cycle track crossings – traffic conflict technique, Journal of Transportation Safety & Security 12: pp. 194–209, 2020. https://doi.org/10.1080/19439962.2019.1622615

[5] Y.A. Ünvan, Market basket analysis with association rules, Communications in Statistics - Theory and Methods, 2020. https://doi.org/10.1080/03610926.2020.1716255

[6] D.T. Larose, C.T. Larose, Discovering Knowledge in Data, 2nd edition, Wiley, 2016.

[7] T. Morzy, Eksploracja danych. Metody i algorytmy, PWN, Warsaw, 2013.

[8] A. Shi, B. Mou, J.C. Correl, Association analysis for oxalate concentration in spinach, Euphytica, 2003. https://doi.org/10.1007/s10681-016-1740-0

[9] M. Lasek, M. Pęczkowski, Analiza asocjacji i reguły asocjacyjne w badaniu wyborów zajęć dydaktycznych dokonywanych przez studentów. Zastosowanie algorytmu Apriori, Ekonomia. Rynek. Gospodarka, Warsaw, 2013.

[10] T. Klimanek, M. Szymkowiak, T. Józefowski, Application of market basket analysis in biological disability, Research Papers of Wrocław University of Economics 507, 2018. https://doi.org/10.15611/pn.2018.507.09

[11] A.M. Ahmed, A.A. Bakar AA, A.R. Hamdana, S.M. Abdullah, O. Jaafarb, Sequential pattern discovery algorithm for Malaysia rainfall prediction. Acta Phys Pol A 2015. http://dx.doi.org/10.12693/APhysPolA.128.B-324

[12] A. Nicał, H. Anysz, The quality management in precast concrete production and delivery processes supported by association analysis, International Journal of Environmental Science and Technology, 2019. https://doi.org/10.1007/s13762-019-02597-9

[13] H. Anysz, A. Foremny, J. Kulejewski, Comparison of ANN classifier to the neuro-fuzzy system for collusion detection in the tender procedures of road construction sector. IOP Conf Ser Mater Sci Eng., 2019. https://dx.doi.org/10.1088/1757-899x/471/11/112064

[14] H. Anysz, B. Buczkowski, The association analysis for risk evaluation of significant delay occurrence in the completion date of construction project, International Journal of Environmental Science and Technology, 2018. https://doi.org/10.1007/s13762-018-1892-7

[15] K. Guerts, G, Wets, T. Brijs, K. Vanhoof, Profiling High Frequency Accident Locations Using Association Rules, Transportation Research Record ï Journal of the Transportation Research Board, 1840, 2003. http://dx.doi.org/10.3141/1840-14
[16] A. Pande, M. Abdel-Aty, Market basket analysis of crash data from large jurisdictions and its potential as a decision support tool, Elsevier, Safety Science 47: pp. 145–154, 2009. https://doi.org/10.1016/j.ssci.2007.12.001
[17] C. Xu, J. Bao, C. Wang, P. Liu, Association rule analysis of factors contributing to extraordinarily severe traffic crashes in China, Journal of Safety Research 67: 65-75, 2018. doi: 10.1016/j.jsr.2018.09.013
[18] D. Nenadić, Ranking dangerous sections of the road using MCDM model. Decision Making: Applications in Management and Engineering, 2(1): pp. 115–131, 2019. Retrieved from https://dmame.rabek.org/index.php/dmame/article/view/31
[19] P. Olszewski, P. Szagała, D. Rabczenko, & A. Zielińska, Investigating safety of vulnerable road users in selected EU countries. Journal of Safety Research, 68: pp. 49–57, 2019. https://doi.org/10.1016/j.jsr.2018.12.001
[20] https://ec.europa.eu/transport/road_safety/specialist/statistics# (access June 2019)

Identyfikacja czynników i warunków prowadzących do poważnych wypadków rowerzystów z wykorzystaniem analizy asocjacji

Słowa kluczowe: wypadki rowerzystów, ruch rowerowy, obrażenia rowerzystów, analiza asocjacji, analiza koszykowa

Streszczenie:

Opublikowane przez Komisję Europejską dane w CARE (Community database on Accidents in the Roads in Europe) wskazują, że Polska jest krajem o najwyższym wskaźniku śmiertelności rowerzystów w efekcie wypadków drogowych w okresie 2009-2013 r. Jednocześnie rower staje się coraz bardziej popularnym środkiem transportu i sposobem rekreacji. W Warszawie ruch rowerowy wzrasta od 1 do 3% rocznie. To zjawisko oraz coraz gęstszy ruch drogowy było w Polsce przyczyną 4233 zarejestrowanych wypadków w 2018 r, w których ucierpieli rowerzyści. W 286 przypadkach wypadek był przyczyną śmierci rowerzysty. W świetle powyższych informacji bardzo istotne staje się wskazanie najważniejszych czynników i warunków drogowych towarzyszących poważnym wypadkom rowerzystów. Statystyki z użyciem jednej czy dwu zmiennych są niewystarczające, gdyż okoliczności towarzyszących wypadkom jest zwykle więcej. Do badania tych okoliczności wykorzystano więc analizę asocjacji zwaną też analizą koszykową.

Obserwowanemu wzrostowi ruchu rowerowego (w Warszawie podwoił się w ciągu ostatnich kilku lat) towarzyszy także wzrost długości wybudowanych ścieżek rowerowych (np. w Warszawie od 305 km w 2011 r. do 645 km w 2019 r.). Niestety podobny wzrost dotyczy liczby wypadków w udziałem rowerzystów, wśród których także są wypadki śmiertelne. Całkowitą liczbę wypadków rowerzystów w Polsce wraz z liczbą wypadków śmiertelnych w latach 2011–2019 pokazano na rys. 3.

Analiza asocjacji to wyszukiwanie związków pomiędzy zdarzeniem (lub jednoczesnym wystąpieniem kilku zdarzeń) towarzyszących innemu, obserwowanemu zdarzeniu, bądź zjawisku. Inna nazwa takich analiz, to analiza koszykowa, która powstała od pierwotnego zastosowania tj. badań zawartości koszyków klientów w sklepach wielkopowierzchniowych. Wtedy celem było zwiększanie sprzedaży, jednak przydatność tego rodzaju analiz została zauważona przez naukowców i są one wykorzystywane np. w biologii [8], edukacji [9]. Naturalnym zastosowaniem analizy asocjacji jest medycyna (np. w [10]). Symptomy choroby stanowią tam poprzednik reguły, a następnikiem jest wystąpienie choroby. Zastosowania analizy koszykowej można także znaleźć w rozwiązaniach problemów z zakresu inżynierii lądowej np. w zagadnieniach dotyczących zarządzania jakością [12], w ocenie ryzyka [14], w ocenie rzetelności ofert w postępowaniach przetargowych [13]. Reguł asocjacyjnych poszukiwano już także w zagadnieniach związanych z bezpieczeństwem ruchu drogowego [15, 16, 17]. Analiza asocjacji posługuje się trzema podstawowymi wskaźnikami: wsparciem (sup od ang. support) oznaczającym udział zdarzeń z jednoczesnym wystąpieniem poprzednika i następnika w całkowitej liczbie zdarzeń, ufnością ($conf$ od ang. confidence) wyrażającym stosunek liczby zdarzeń z jednoczesnym wystąpieniem poprzednika i następnika do liczby zdarzeń, w których wystąpił dany poprzednik, przyrostem (ang. $lift$), który zdefiniowany jest jako iloraz ufności i prawdopodobieństwa wystąpienia następnika. Znalezione reguły o przyroście mniejszym niż 1 nie wyjaśniają zjawiska. Przykładowe obliczenia tych trzech parametrów przedstawiono w tab. 1.

Analizie poddano bazę danych stworzoną przez Instytut Transportu Samochodowego na podstawie bazy [20]. Zawiera ona informacje o 21470 wypadkach rowerzystów w Polsce jakie miały miejsce w latach 2009-2013. Rodzaje informacji jakie zawiera każdy z rekordów w bazie to: stopień obrażeń rowerzysty (lekkie, ciężkie lub śmiertelne), miejsce zdarzenia (obszar miejski lub poza nim), rodzaj ulicy/drogi (jedno lub dwujezdniowa, jedno lub dwukierunkowa), lokalizacja zdarzenia na skrzyżowaniu lub poza nim, czy skrzyżowanie było wyposażone w sygnalizację świetlną, jaki był stopień oświetlenie naturalnego (dzień, zmierzch, noc). Bazę danych (której fragment przedstawiono na rys. 4) przekształcono do postaci z binarnymi wartościami (patrz tab. 2). Niezależnie od postaci przedstawianych informacji (tabelarycznie lub graficznie) analiza więcej niż dwóch czynników jednocześnie wpływających na stopień obrażeń jest albo trudna, albo informacje są nieczytelne. Podstawowe statystyki dotyczące wypadków rowerowych przedstawiono w rozdziale 3.1. Ze względu na znaczne różnice w strukturze wypadków w obszarze miejskim i poza nim, zdecydowano na poszukiwanie reguł asocjacyjnych odrębnie dla tych dwóch obszarów. Drugim założeniem było połączenie kategorii obrażeń ciężkich i śmiertelnych w jedną kategorię oznaczaną FS (ang. fatal and serious injuries), gdyż poszukiwanie reguł wyłącznie dla wypadków śmiertelnych pozwoliło znaleźć małą liczbę mało znaczących reguł (z przyrostem poniżej 1). Najsilniejsze, istotne reguły znalezione po przyjęciu wyżej opisanych założeń przedstawiono w tab. 6, 7, 8 i 9. Regułą o największej ufności jest następująca reguła (nr 10 w tab. 9): jeśli wypadek miał miejsce poza terenem miejskim na skrzyżowaniu drogi jednojezdniowej, dwukierunkowej, w czasie zmierzchu lub nocą, to w wyniku wypadku rowerzysta zmarł lub doznał ciężkich obrażeń ciała, z ufnością 80,0%. Nie znaleziono natomiast istotnych reguł dotyczących wypadków ze skutkiem FS dla obszarów miejskich. Najbardziej zaskakującym wnioskiem z przeprowadzonych analiz jest mała istotność sygnalizacji świetlnej w obszarze miejskim dla ciężkich wypadków w ruchu rowerowym. Natomiast poza obszarem miejskim, o ile skrzyżowanie posiada sygnalizację świetlną, w wypadku z udziałem rowerzysty można spodziewać się jego lekkich obrażeń z ufnością 65,2% przy wsparciu reguły 41,4%. Należy podkreślić, że przedstawione w artykule analizy były wykonane w oparciu o bazę danych, która nie zawiera wielu informacji istotnych z punktu widzenia przyczyn wypadków drogowych z udziałem rowerzystów takich jak: sztuczne oświetlenie ulicy lub drogi, oświetlenie osobiste rowerzysty, czy wypadek miał miejsce na ścieżce rowerowej, jakie były warunki pogodowe. Uwzględnienie ich w analizie koszykowej mogłoby doprowadzić do znalezienia reguł o większej istotności oraz do wniosków, które jeszcze w większym stopniu mogłyby się przyczynić do zapobiegania wypadkom rowerzystów.

Received: 2020-11-25, Revised: 2021-01-22