Trauma Mon. 2018 May; 23(3):e57926. doi: 10.5812/traumamon.57926.

Published online 2017 December 8. **Research Article**

Application of Capture-Recapture Method to Estimate Traffic Accident Mortality Rate

Narges Entezami,[1] Seyed-Saeed Hashemi-Nazari,[2,*] Hamid Soori,[2] Ardeshir Khosravi,[3] and Hamid Heidarian Miri[4]

[1]Department of Epidemiology, School of Public Health, Shahid Beheshti University of Medical Sciences, Tehran, IR Iran
[2]Department of Epidemiology, Safety Promotion and Injury Prevention Research Center, School of Public Health, Shahid Beheshti University of Medical Sciences, Tehran, IR Iran
[3]Iranian Ministry of Health and Medical Education and Research Center for Aging, University of Social Welfare and Rehabilitation, Tehran, IR Iran
[4]Assistant Professor, Social Determinants of Health Research Center, Mashhad University of Medical Sciences, Mashhad, IR Iran

[*]*Corresponding author*: Seyed-Saeed Hashemi-Nazari, Department of Epidemiology, Safety Promotion and Injury Prevention Research Center, School of Public Health, Shahid Beheshti University of Medical Sciences, Tehran, IR Iran. Tel/Fax: +98-2122439980, E-mail: saeedh_1999@yahoo.com

Received 2017 July 10; **Revised** 2017 July 15; **Accepted** 2017 September 06.

Abstract

Background: In many countries, disease registries are the main source of information necessary for planning and monitoring health-improving programs. In Iran, there is no unified registry system for deaths due to traffic accidents, so organizations that are involved in traffic accidents report different statistics. This study aimed to assess traffic accident mortality rate and completeness of death registries in Northern provinces of Iran from March 21, 2010 to March 20, 2011 using capture-recapture method.
Methods: Data of all traffic accident victims including last name, age, sex, and time of death confirmed by traffic police legal forensics organization, and hospitals were used to collect. Then, different log-linear models proposed in the context of capture-recapture method were fitted to predict the number of cases missed by the sources. Akaike information criterion (AIC) and Bayesian information criterion (BIC) were used as goodness of fit indices to identify preferred fitted models.
Results: According to the prediction of the most fitted model, the total number of traffic accident deaths was 3857 of which, 727 cases were not documented by any of the responsible organizations. Mortality rate due to traffic accidents was estimated as 53.73 with 95% confidence interval (52.41 - 55.15) per 100000 population of the three provinces. The rates of legal medicine organization, hospitals, and traffic police department were 65.95%, 36.06%, and 16.85%, respectively.
Conclusions: Incidence rate of fatal traffic accidents in Northern provinces of Iran is higher in comparison with other provinces. It seems that statistics reported by responsible organizations are underestimated, hence the need for application of methods like the capture-recapture method.

Keywords: Mortality, Capture-Recapture, Road Traffic Injuries

1. Background

Road traffic injury (RTI) imposes great burden of mortality and morbidity on families and societies in Iran. Therefore, it is the first priority to many people and organizations to reduce RTI, all of whom need accurate information for successful planning and evaluation of preventive programs (1). However, in Iran, in spite of the high incidence of RTI, there is no specific registry or surveillance system to gather and report information on RTI, and involved organizations usually report different and, more importantly, underestimated rates for RTI.

One popular technique to deal with this problem is capture-recapture method. Capture-recapture relies on the degree of overlap among two or more incomplete lists to estimate the number of cases missed from all lists and consequently the total number of cases (2). It was originally used in ecology to determine the number of animals in a specified area (3, 4). Recently, these methods have widely been used to estimate incidence and prevalence of health related outcomes (5-8) and to evaluate the sensitivity of surveillance or registry systems (9-11).

There are four assumptions on which this method is based: first, the population should be closed that means no change should occur in the size or structure of the population during study time; second, subjects should have equal probability of being captured in given sources; third, sources should act independently that means being captured by one source should not affect the probability of being captured by other sources; and four, there must be

sufficient information to make a perfect linkage among sources (2, 12, 13).

Northern provinces of Iran (Guilan, Golestan, and Mazandaran) are among the most accident-prone areas of the country as they attract many tourists for their geographical and natural wonders. The purpose of this study was to estimate the mortality rate due to road traffic accidents and the level of completeness of death registries in legal medicine organization (LMO), traffic police department (TPD), and hospitals in Northern provinces of Iran.

2. Methods

2.1. Design and Population Study

This study is a descriptive cross-sectional study in which capture-recapture method was used to estimate fatal RTI rate. Population under study was those who were killed in traffic accidents in Northern provinces of Iran including Golestan, Mazandaran, and Guilan from March 21, 2010 to March 20, 2011.

2.2. Data Sources

We took data from three organizations as LMO, hospitals, and TPD. LMO records information of any death due to road traffic accidents. Hospitals record information on injured people who are taken to a hospital and die during the 30 following days. TPD record RTI deaths that happen at the accident scene.

2.3. Analysis

Statistical analysis of this study was conducted to estimate the number of fatal traffic accidents and consequently the corresponding incidence rate as well as the completeness of reports made by relevant organizations.

To estimate the total number of cases, the first step was to count the cases that existed uniquely in each source besides those that were common between any combinations of two sources as well as cases seen in all three sources. For this purpose, linking variables were first name, surname, age, sex, and date of death. At least four out of five of these criteria had to be met for records to be regarded as one unique case.

2.3.1. Tow Source Capture- Recapture

We calculated chapman estimations of the total number of cases of fatal RTI by pair wise matching of the three sources. Suppose N_A is the number of cases by source A, N_B is the number of cases by source B, and N_{AB} is the number of cases by both sources, then chapman estimate of the number of cases and the corresponding variance (Var) and confidence limits (95% CI) are computed as follows (14, 15):

$$N = \frac{(N_A + 1)(N_B + 1)}{(N_{AB} + 1)} - 1 \quad (1)$$

$$Var = \frac{(N_A + 1)(N_B + 1)(N_A - N_{AB})(N_B - N_{AB})}{(N_{AB} + 1)^2 (N_{AB} + 2)} \quad (2)$$

$$95\%\ \mathrm{CI} = \mathrm{N} \pm 1.96\sqrt{\mathrm{Var\,(N)}} \quad (3)$$

2.3.2. Three-Source Capture-Recapture

With multiple registries, a more powerful form of capture-recapture approach can be done through log-linear models to adjust for possible dependency and heterogeneity (13, 16). In this context, the number of registered cases form the dependent variable and each registry serves as an independent variable that takes zero if the cases were missed from that and one otherwise. Therefore, with n registries there would be 2^n rows in the data set in which the value of dependent variable when all n independent variables are zero is missing. Then, the fitted log-linear model can be used to predict this missing value.

With three sources and based on whether or not dependency exists among sources, all possible combination of main effects and their interaction form eight models that differ in terms of type and number of product terms. The most general model is:

m_{ijk}: expected frequency for cell ijk

U, common parameter

U_1, U_2 and U_3, main effect (first order effect) parameters

U_{12}, U_{13} and U_{23}, second order effect (two-way interaction) parameters

U_{123}, the highest order effect (three-way interaction) parameter

In formula 1, there are 8 parameters whereas the number of categories that provide information is seven. Therefore, a concrete presumption is that three-way interaction parameter is zero (16-18). Akaike information criterion (AIC) and Bayesian information criterion (BIC) were used as a basis for judgment about the performance of models.

Stratified analysis was performed over sex and age groups. Age groups were defined as 0 to 24, 25 to 44, 45 to 64, and 65 or over. If there was no observation in one cell, correction coefficient $1/2^{k-1}$, k stands for the number of sources, will be added to all cells in order to prevent spars data bias (19, 20). However, in this study, all sex-age cells contained observations.

The completeness of each source was calculated by dividing the number of fatal traffic accidents reported by that source by total number of fatal traffic accident estimated by the selective log linear model. It is worthy to add

$$\log(m_{ijk}) = U + U_{1(i)} + U_{2(j)} + U_{3(k)} + U_{12(ij)} + U_{13(ik)} + U_{23(jk)} + U_{123(ijk)} \quad (4)$$

that the identification of a number of victims (1 in hospitals, 50 TPD, and 9 LMO, respectively) was impossible because their names and surnames were unavailable in the reported list. For data analysis, we used R-capture package in R software.

To calculate the incidence rate, the numbers of population in the three given provinces, reported by Iran's Statistics Center, were used as denominator.

3. Results

Totally, 4525 traffic deaths were recorded by Hospitals (1390), TPD (600), and LMO (2535) in Iran's Northern provinces from March 21, 2010 to March 20, 2011. The distribution of recorded cases over these three sources is shown in Table 1. As can be seen, of all the recorded cases, 1899 were captured just by one of the three sources and the minimum number of common cases was observed between hospitals and TPD. Taking a descriptive look, we observed that mean of age was 40.01 ± 0.37 and 56% of the victims were in age group 15 to 44. In addition, sex ratio (men to women) was 4 indicating men constituted about 80 percent of victims.

Table 1. Distribution of Number of Traffic Deaths Captured by Data Sources

Data Source			Observed Data
LMO	**H**	**TPD**	
Yes[b]	No[b]	No	1311
No	Yes	No	446
No	No	Yes	142
Yes	Yes	No	773
Yes	No	Yes	287
No	Yes	Yes	7
Yes	Yes	Yes	164
No	No	No	Unobserved

Abbreviations: H, Hospitals; LMO, legal Medicine Organization; TPD, Traffic Police Department.
[b]Yes, indicates presence; No indicates absence of the data sources.

Results obtained from chapman estimator showed that the estimates for TPD and hospital sources are different from those of two others. Stratified analysis revealed the existence of almost similar pattern in sex and age groups (Table 2).

As described in the previous section, log-linear model was used to estimate the total number of deaths. To adjust for dependencies between sources, double (two-component) product terms were entered in Log-linear models.

With three sources, there would be eight possible models. Model 1 simply includes three sources without any product term (LMO, H,TPD), models 2 to 4 contain one double product term (LMO × H,TPD; LMO × TPD, H; and H × TPD, LMO), the remaining models had two double product terms with one source in common (LMO × H, H × TPD; LMO × TPD, H × TPD; and LMO × H, LMO × TPD), and finally one model with three double product terms in which each two-way combination has one source in common (LMO × H, LMO × TPD, H × TPD). All the fitted models and their corresponding properties are shown in Table 3. In addition, Table 4 shows the estimated number of death for each age and sex groups.

As Table 3 shows, models 7 and 8 are marked with caution and thus, they were excluded from further consideration. Among six remaining models, model 4, based on AIC and BIC, best fitted the data. This model estimated that there were 727 deaths that have not been captured by the regarding sources. The estimate of total number of deaths was 3857. The incidence rate of fatal traffic accident in Northern provinces of Iran was 53.73 per 100000 with 95% confidence interval (52.41 - 55.15). Completeness rates of LMO, Hospitals, and TPD were 65.95%, 36.06%, and 16.85%, respectively.

4. Discussion

Capture-recapture method is regarded as an approach to estimate the prevalence of an event based on degrees to which two or more lists overlap (2). In the present study, three lists made by LMO, hospitals, and TPD were used to estimate the incidence rate of traffic accident deaths in Northern provinces of Iran from March 21, 2010 to March 20, 2011. The results show that the number of traffic accident deaths in the mentioned time and place could have been 3857 with 95% CI (3762 - 3959) that comes to an incidence rate of 53.73 with 95% CI (52.41 - 55.15) per 100000. The corresponding rates (per 100000) in different provinces of Iran were 28.8 in Khuzestan, 51.3 in Kermanshah (21, 22), and 21.07 in Isfahan (23). Besides, a study showed that fatal traffic accident rate was 23.19 per 100000 in Mazandaran during 2007 to 2010 (24). All the mentioned rates are less than what estimated in this study. Use of three sources of information and application of capture-recapture method help to incorporate into the estimated

Table 2. Estimated Total Number (95% confidence interval) of Fatal Road Traffic Injuries Among Different Subgroups by Using 2 Source Capture-Recapture Analysis, Northern Provinces of Iran, During 2010 to 2011

Sex-Age Group	LMO/H	LMO/TPD	H/TPD
	Unstratified analysis		
All	3760 (3651-3869)	3371 (3231-3511)	4859 (4286-5432)
	stratified analysis		
Men	3000 (2905-3095)	2748 (2617-2879)	3874 (3372-4376)
Women	760 (706-814)	624 (573-675)	969 (704-1234)
Total	3760 (3611-3909)	3372 (3190-3554)	4843 (4076-5610)
≥ 24	1069 (1015-1123)	914 (860-968)	1202 (959-1445)
25 - 44	1279 (1209-1349)	1182 (1094-1270)	1973 (1557-2389)
45 - 64	756 (709-803)	715 (643-787)	1077 (780-1374)
≥ 65	653 (610-696)	534 (479-589)	577 (441-713)
Total	3757 (3649-3865)	3345 (3206-3484)	4829 (4274-5384)

Abbreviations: H, Hospitals; LMO, legal Medicine Organization; TPD, Traffic Police Department.

Table 3. Estimated Total Number of Fatal Road Traffic Injuries According to Log-Linear Models, Northern Provinces of Iran During 2010 to 2011

Model NO.	Log-linear model	X[b]	N[c]	95% CI for N	df	BIC	AIC	Information fit
1	LMO, H, TPD	724	3854	3756 - 3957	2	203	172.7	OK
2	LMO × H, TPD[a,c]	849	3979	3807 - 4175	1	207.7	171.4	OK
3	LMO × TPD, H	831	3961	3846 - 4085	1	183.3	147	OK
4	H × TPD, LMO	727	3857	3762 - 3959	1	100.4	64.1	OK
5	LMO × H, H × TPD	649	3779	3644 - 3932	0	106.8	64.4	OK
6	LMO × TPD, H × TPD	756	3886	3779 - 4002	0	106.8	64.4	OK
7	LMO × H, LMO × TPD	9047	12177	7584 - 24139	1	98.8	62.5	Warning[d]
8	LMO × H, LMO × TPD, H × TPD	8768	11898	indefinite	0	106.8	64.4	Warning[e]

Abbreviations: AIC, Akaike Information Criterion; BIC, Bayesian Information Criterion; CI, Confidence Interval; df, Degrees of Freedom; H, Hospitals; LMO, legal Medicine Organization; TPD, Traffic Police Department.
[a] LMO × H, TPD denotes interaction between LMO and hospitals, etc.
[b] Estimate of number of deaths that are not captured by any source.
[c] Estimate of total number of deaths.
[d] Model fitness is questionable (algorithm did not converge, non-positive sigma estimate for a normal heterogeneous model or large asymptotic bias).
[e] Design matrix is not of full rank; so some model's coefficients are not estimable.

cases that had been missed from reports. On the other hand, heavy traffic, climatic states, and condition of roads in Northern provinces might have an important role in the increased number of traffic accidents.

Regarding percent completeness, the highest value belonged to LMO (65.95) followed by hospitals (36.06), and TPD (16.85) that is consistent with the results of other studies (25, 26). The low completeness of TPD could be well explained by the fact that it captures those deaths that happen immediately in a traffic accident so they miss a large proportion of cases occurring in hospitals. However, the highest completeness goes with police in Utopia, since the police follow the victims for a month after the accident (27).

Estimates based on capture-recapture method are more trustable under following four assumptions (2, 12, 13). First, the population should be closed that means no change should occur in the size or structure of the population during study time. In the present study, this assumption was almost met since the population was traffic accident deaths not the residents.

Second, subjects should have equal probability of being captured by a certain source. Stratified analysis showed that the mentioned probability did not depend on sex and age, as the estimate did not show a considerable change after stratification. Assessment of intra source heterogeneity needs to extend the stratification further over other variables like role of victim (driver, passenger,

Table 4. Estimated Total Number of Fatal Road Traffic Injuries According to Log-Linear Models, by Age and Sex, Northern Provinces of Iran During 2010 to 2011

Sex-Age Group	Minimum AIC	BIC	X[a]	N[b]	95% CI for N
Men	61.2	96.24	581	3100	3016 - 3192
Female	52.7	83.6	168	779	729 - 839
All			749	3879	
≥ 24	55.3	89	190	1091	1040 - 1150
25 - 44	55.3	85.1	275	1332	1274 - 1397
45 - 64	51.7	78.4	144	781	741 - 828
≥ 64	49.6	75.2	135	670	630 - 718
All			744	3874	

Abbreviations: AIC, Akaike Information Criterion; BIC, Bayesian Information Criterion.
[a] Estimate of number of deaths that are not captured by any source.
[b] Estimate of total number of deaths.

or pedestrian) or the place of death. Because of limited access to the required information, heterogeneity assessment was restricted to just sex and age.

Third, sources should work independently, meaning that the probability of being captured by a certain source should not depend on being captured by another source. Two-source analysis showed that the estimates derived from hospitals-TPD were considerably higher than the estimates of LMO-H and LMO-TPD, indicating the existence of a negative association between hospitals and TPD and a positive association between LMO and hospitals as well as between LMO and TPD. The violation of this assumption was not an issue since in log linear models, interaction terms account for possible independence between the sources.

Fourth, there must be sufficient information to make a perfect linkage among sources. Fortunately, this task was satisfactorily accomplished using variables of name, surname, age, sex, and date of death.

4.1. Conclusion

The absence of a valid and reliable system to register traffic accident victims is more problematic when the responsible organizations report different statistics that are all biased. In this situation, capture-recapture method could be regarded as a powerful tool to validly estimate the number of cases that might have happened.

Acknowledgments

The authors wish to thank the officials of Iran's ministry of health, forensic medicine organization, and traffic police for their support in making the data accessible.

References

1. Mohan D. *Road traffic injury prevention training manual*. World Health Organization; 2006.
2. Hook EB, Hsia MS, Regal RR. Accuracy of capture-recapture estimates of prevalence. *Epidemiol Methods*. 2012;**1**(1):1–11. doi: 10.1515/2161-962x.1000.
3. Pollock KH. Modeling capture, recapture, and removal statistics for estimation of demographic parameters for fish and wildlife populations: Past, present, and future. *J Am Stat Assoc*. 1991;**86**(413):225–38. doi: 10.2307/2289733.
4. Seber GAF. A Review of estimating animal abundance II. *Int Stat Rev*. 1992;**60**(2):129. doi: 10.2307/1403646.
5. Gimenez-Duran J, Galmes-Truyols A, Gonzalez-Cortijo T, Bosch-Isabel C, Nicolau-Riutort A, de Mateo-Ontanon S. Estimated incidence of tuberculosis by capture-recapture in the Balearic Islands, 2010-2012. *Int J Tuberc Lung Dis*. 2014;**18**(11):1357–62. doi: 10.5588/ijtld.14.0251. [PubMed: 25299871].
6. Hernandez-Martin A, Garcia-Doval I, Aranegui B, de Unamuno P, Rodriguez-Pazos L, Gonzalez-Ensenat MA, et al. Prevalence of autosomal recessive congenital ichthyosis: a population-based study using the capture-recapture method in Spain. *J Am Acad Dermatol*. 2012;**67**(2):240–4. doi: 10.1016/j.jaad.2011.07.033. [PubMed: 22000705].
7. Preux PM, Druet-Cabanac M, Couratier P, Debrock C, Truong T, Marcharia W, et al. Estimation of the amyotrophic lateral sclerosis incidence by capture-recapture method in the Limousin region of France. *J Clin Epidemiol*. 2000;**53**(10):1025–9. doi: 10.1016/S0895-4356(00)00212-2. [PubMed: 11027935].
8. Tilling K, Sterne JA, Wolfe CD. Estimation of the incidence of stroke using a capture-recapture model including covariates. *Int J Epidemiol*. 2001;**30**(6):1351–9. discussion 1359-60. doi: 10.1093/ije/30.6.1351. [PubMed: 11821345].
9. Jansson A, Arneborn M, Ekdahl K. Sensitivity of the Swedish statutory surveillance system for communicable diseases 1998-2002, assessed by the capture-recapture method. *Epidemiol Infect*. 2005;**133**(3):401–7. doi: 10.1017/S0950268804003632. [PubMed: 15962546].
10. Roberts I, Scragg R. Application of capture-recapture methodology to estimate the completeness of child injury surveillance. *J Paediatr Child Health*. 1994;**30**(6):513–4. doi: 10.1111/j.1440-1754.1994.tb00723.x. [PubMed: 7865264].
11. Robles SC, Marrett LD, Clarke EA, Risch HA. An application of capture-recapture methods to the estimation of completeness of cancer registration. *J Clin Epidemiol*. 1988;**41**(5):495–501. doi: 10.1016/0895-4356(88)90052-2. [PubMed: 3367181].
12. Capture-recapture and multiple-record systems estimation I: History and theoretical development. International Working Group for Disease Monitoring and Forecasting. *Am J Epidemiol*. 1995;**142**(10):1047–58. [PubMed: 7485050].

13. Hook EB, Regal RR. Capture-recapture methods in epidemiology: methods and limitations. *Epidemiol Rev*. 1995;**17**(2):243–64. doi: 10.1093/oxfordjournals.epirev.a036192. [PubMed: 8654510].
14. Chapman DG. *Some properties of the hypergeometric distribution with applications to zoological sample censuses*. University of California Press; 1951.
15. Seber GAF. The Effects of Trap Response on Tag Recapture Estimates. *Biometrics*. 1970;**26**(1):13. doi: 10.2307/2529040.
16. Van Hest R. *Capture-recapture methods in surveillance of tuberculosis and other infectious diseases*. Erasmus University Rotterdam; 2007.
17. Chao A. An overview of closed capture-recapture models. *J Agric Biol Environ Stat*. 2001;**6**(2):158–75. doi: 10.1198/108571101750524670.
18. Chao A, Tsay PK, Lin SH, Shau WY, Chao DY. The applications of capture-recapture models to epidemiological data. *Stat Med*. 2001;**20**(20):3123–57. doi: 10.1002/sim.996. [PubMed: 11590637].
19. Evans MA, Bonett DG. Bias reduction for multiple-recapture estimators of closed population size. *Biometrics*. 1994;**50**(2):388–95. doi: 10.2307/2533382. [PubMed: 8068839].
20. Hook EB, Regal RR. Accuracy of alternative approaches to capture-recapture estimates of disease frequency: internal validity analysis of data from five sources. *Am J Epidemiol*. 2000;**152**(8):771–9. doi: 10.1093/aje/152.8.771. [PubMed: 11052556].
21. Hashemi Nazari SS, Kazemian M, Hosseini F. Trend of five years traffic accident mortality in Khuzestan province (2006-2010) [In Persian]. *Sci J Forensic Med*. 2011;**17**(2):123–9.
22. Izadi N, Najafi F, Khosravi A. Estimation of mortality and calculated years of lost life from road traffic injuries [In Persian]. *J Mazandaran Univ Med Sci*. 2014;**24**(112):51–8.
23. Maracy MR, Tabar Isfahani M . The burden of road traffic injuries in Isfahan, Iran in 2010 [In Persian]. *J Kerman Univ Med Sci*. 2013;**20**(5):505–19.
24. Yazdani Cherati J, Ahmadi Baseri E, Ghadami M. Mapping of mortality rate in suburban accidents, Mazandaran Pprovince, 2007-2010 [In Persian]. *J Mazandaran Univ Med Sci*. 2013;**22**(97):50–8.
25. Khorasani-Zavareh D, Bigdeli M, Hatami H, Meshkini A, Mohammadi R. Application of mark-recapture to evaluate preventive road traffic injury policy. *J Inj Violence Res*. 2014;**6**(2):97–8. doi: 10.5249/jivr.v6i2.479. [PubMed: 24284811].
26. Motevalian SA, Mahmoodi M, Majdzadeh R. Estimation of death due to road traffic injuries in Kerman district: application of capture-recapture method [In Persian]. *J Sch Public Health Inst Public Health Res*. 2007;**5**(2):61–72.
27. Abegaz T, Berhane Y, Worku A, Assrat A, Assefa A. Road traffic deaths and injuries are under-reported in Ethiopia: a capture-recapture method. *PLoS One*. 2014;**9**(7). e103001. doi: 10.1371/journal.pone.0103001. [PubMed: 25054440].