

Title of study: Factors that affect road safety in the North West Province of South Africa

Objective of study

The main objective of this study is to identify key risk factors for road accidents in the North West Province in which humans are the victim. The study aims to answer each of the following research questions:

- What are the factors that affect road accidents involving humans in the North West province?
- What is the perception of ordinary people in the North West Province as to what causes road accidents?
- What suitable remedial actions could be taken by the Road Authorities in order to reduce the incidence of road accidents?

Study design

The design of the study is cross-sectional, descriptive and household based. This is different from longitudinal study designs in which repeated observations are made. In this study, data is gathered from respondents only once.

List of variables of study

The study is based on a total of 65 variables of study. These variables are related to road accidents in one way or another. Most of the 65 variables are socio-economic, demographic and transport related questions that are helpful in explaining the occurrence of road accidents in the North West Province.

The outcome (dependent) variable of study measures the infliction of an accident on a human by a vehicle on the road (**carhuman**). Examples of independent variables of study that were used in this study were: nature of road, poor street lighting, condition of brakes, condition of vehicle lights, road worthiness of vehicle, overloading of passengers, overloading of goods, attitude of driver toward other road users, consumption of alcohol while driving, adhering to traffic rules while driving, driving skills, availability of traffic officers for enforcing traffic regulations, use of roadblocks for enforcing traffic regulations, use of speed traps for enforcing traffic regulations, use of traffic calming procedures for ensuring road safety, and observing road signs effectively.

Sampling frame and technique

The study is based on a random sample of **n=1161** respondents identified from 5 regions of the North West Province based on a **three-stage cluster sampling** procedure in which sampling weights were used. The province was divided into **5 regions** and **54 primary sampling units or enumeration areas**. Each of the 5 regions is made up of several place names.

Stage 1 of the selection procedure involved the selection of regions (Central, East, North, South and West).

Stage 2 of the selection procedure involved the selection of primary sampling units or enumeration areas.

Stage 3 of the selection procedure involved the selection of an eligible respondent from households.

At each household, one respondent was selected at random for interview. As such, the probability of selection at Stage 3 of the selection procedure was 100% or 1. Ideally, the selection of eligible respondents at the household level should be randomized. However, this was not done due to resource constraints such as time and unavailability of heads of households due to work obligations at the time of the interview. Also, only 19 households were selected from the Northern part of the North West Province for the study. The fact that the probability of selection at Stage 3 was 100% for all respondents who took part in the study, and the fact that only 19 households were selected from the Northern part of the province constitutes a minor limitation of study.

Table 1: Selection of households per region

Region	Number of households	Number of respondents
Central	308	308
East	503	503
North	19	19
South	176	176
West	155	155
Total	1, 161	1, 161

Selection of households or dwelling units

The listing exercise provided a complete list of dwelling units (DU) in the respective selected PSUs. All the dwelling units were considered for selection whether vacant or occupied.

The selection of the sub-sample was based on a fixed number of 10% DUs in each place name. A constraint was imposed on small places where the total dwelling units are equal or less than 10, all the dwellings units are taken. Where there are multiple households in a dwelling unit, all the households are enumerated. Replacement of refusals, vacant dwellings or non-contact was not considered because of their impact on the probability of selection. Concerted efforts were made to improve the response rates.

Inclusion criteria

All members of the household aged 18 years and above were considered for selection as respondents. The respondent was selected randomly using a grid. The use of a fixed sample when selecting dwelling units implies that the workload is distributed equally for the fieldworkers. However, some dwelling units carried more weights than others. It is not advisable to replace vacant

dwelling units, refusals or non-contacts in view of the distortion of the probabilistic nature of the sample design.

The calculation of sampling weights

Strata are regions (1, 2, 3, 4, 5)

PSUs are EAs or areas

Stage 1 (region level) selection probability = $\frac{nm_i}{M} = P_{1i}$ where

n = Number of PSUs (areas) selected from stratum (region) = 5

m_i = Estimated number of respondents in PSU (area) i where $i = 1, \dots, 5$

M = Total number of respondents in all 5 PSUs (areas)

Stage 2 (area level) selection probability =

$$\frac{\text{number of households selected}}{\text{total number of households in area}} = P_{2i}$$

Stage 3 (household level) selection probability =

$$\frac{\text{number of interviewed respondents in selected household}}{\text{total number of respondents in selected household}} = P_{3i} = 1 \text{ since 1 person is}$$

interviewed at each selected household.

Overall selection probability = $P_1 \times P_2 \times P_3 = P$

Sampling weight = $\frac{1}{P} = W$

Recommended adjustment of weights

Adjustment for non-response is not needed as all eligible respondents have been interviewed at each household.

```
. use c:\bos\sello3.dta, clear
```

```
. list id region area weights carhuman curve in 1/60
```

	id	region	area	weights	carhuman	curve
1.	1	central	MADIBOGO	9.5	1	0
2.	2	central	MADIBOGO	9.5	1	1
3.	3	central	MADIBOGO	9.5	0	0
4.	4	central	MADIBOGO	9.5	0	0
5.	5	central	RIETPAN	9.54	0	0
6.	6	central	MADIBOGO	9.5	1	1
7.	7	central	MADIBOGO	9.5	1	0
8.	8	central	MADIBOGO	9.5	0	0
9.	9	central	MADIBOGO	9.5	1	0
10.	10	central	MADIBOGO	9.5	0	1
11.	11	central	MADIBOGO	9.5	1	1
12.	12	central	MADIBOGO	9.5	0	0
13.	13	central	MADIBOGO	9.5	0	0
14.	14	central	MADIBOGO	9.5	0	1
15.	15	central	MADIBOGO	9.5	0	1
16.	16	central	MADIBOGO	9.5	1	0
17.	17	central	MADIBOGO	9.5	0	0
18.	18	central	MADIBOGO	9.5	1	1
19.	19	central	MADIBOGO	9.5	0	1
20.	20	central	MADIBOGO	9.5	1	0
21.	21	central	RIETPAN	9.54	0	0
22.	22	central	MADIBOGO	9.5	1	0
23.	23	central	MADIBOGO	9.5	1	1
24.	24	central	MMABATHO	20.94	1	1
25.	25	central	MADIBOGO	9.5	1	0
26.	26	central	MMABATHO	20.94	1	1
27.	27	central	MADIBOGO	9.5	0	1
28.	28	central	MADIBOGO	9.5	1	1
29.	29	central	MADIBOGO	9.5	1	1
30.	30	central	MADIBOGO	9.5	1	0
31.	31	central	MADIBOGO	9.5	0	1
32.	32	central	MADIBOGO	9.5	0	0
33.	33	central	MADIBOGO	9.5	0	0
34.	34	central	MADIBOGO	9.5	0	0
35.	35	central	MADIBOGO	9.5	1	0
36.	36	central	MADIBOGO	9.5	0	0
37.	37	central	MADIBOGO	9.5	1	0
38.	38	central	MADIBOGO	9.5	1	0
39.	39	central	MADIBOGO	9.5	0	0
40.	40	central	MADIBOGO	9.5	0	0
41.	41	central	MADIBOGO	9.5	0	0
42.	42	central	MADIBOGO	9.5	0	0
43.	43	central	MADIBOGO	9.5	1	1
44.	44	central	MADIBOGO	9.5	1	1
45.	45	central	MMABATHO	20.94	1	0
46.	46	central	MADIBOGO	9.5	1	1
47.	47	central	MADIBOGO	9.5	0	1
48.	48	central	MADIBOGO	9.5	0	0
49.	49	central	MADIBOGO	9.5	1	1
50.	50	central	MADIBOGO	9.5	1	0
51.	51	central	MADIBOGO	9.5	0	1
52.	52	central	MMABATHO	20.94	0	1
53.	53	central	MMABATHO	20.94	0	1
54.	54	central	MMABATHO	20.94	0	0
55.	55	central	MMABATHO	20.94	0	1
56.	56	central	MMABATHO	20.94	1	0
57.	57	central	MMABATHO	20.94	0	1

58.	58	central	MMABATHO	20.94	1	1
59.	59	central	MADIBOGO	9.5	1	0
60.	60	central	MMABATHO	20.94	0	1

```
. list id region area weights carhuman curve in 1150/1161
```

	id	region	area	weights	carhuman	curve
1150.	1150	east	MMAKAUNYANA	1.05	0	1
1151.	1151	east	MMAKAUNYANA	1.05	0	1
1152.	1152	east	MMAKAUNYANA	1.05	1	0
1153.	1153	east	MMAKAUNYANA	1.05	1	0
1154.	1154	east	MMAKAUNYANA	1.05	0	0
1155.	1155	east	MMAKAUNYANA	1.05	0	0
1156.	1156	east	MMAKAUNYANA	1.05	0	0
1157.	1157	east	MMAKAUNYANA	1.05	0	0
1158.	1158	east	WONDERKOP	1871.37	0	1
1159.	1159	east	MMAKAUNYANA	1.05	0	0
1160.	1160	east	MMAKAUNYANA	1.05	0	0
1161.	1161	east	MMAKAUNYANA	1.05	0	0

Statistical methods of data analysis

The **Analysis of Complex Survey Data Sets** was performed in STATA software. Weighted analysis was performed based on the following statistical procedures:

- (1) Frequency tables for discrete variables of study
- (2) Pearson's chi-square tests of association
- (3) Survey binary logistic regression analysis. Unadjusted and adjusted odds ratios were obtained for influential risk factors for road accidents.

The statistical package STATA version 10 was used for the analysis of data.

Reference for the calculation of sampling weights

Calculation of sampling weights by Pierre Foy (**Statistics Canada**)

Pages 71 to 79

Website address: <http://timss.bc.edu/timss199si>

Analysis of Complex Survey Data Sets in STATA version 11

```
. use c:\students\sello3.dta, clear
```

```
. des
```

Contains data from c:\andy\sello3.dta

```
obs:      1,161
vars:      22                               6 Aug 2010 08:07
size:      62,694 (99.9% of memory free)
```

variable name	storage type	display format	value label	variable label
id	int	%8.0g		
region	str7	%7s		
area	str16	%16s		
weights	float	%9.0g		
carhuman	byte	%8.0g		
curve	byte	%8.0g		
lighting	byte	%8.0g		
brakes	byte	%8.0g		
lights	byte	%8.0g		
unworthy	byte	%8.0g		
overload1	byte	%8.0g		
overload2	byte	%8.0g		
attitude	byte	%8.0g		
alcohol	byte	%8.0g		
rules	byte	%8.0g		
skills	byte	%8.0g		
officers	byte	%8.0g		
roadblock	byte	%8.0g		
speedtrap	byte	%8.0g		
calming	byte	%8.0g		
roadsign	byte	%8.0g		
agecar	float	%8.0g		

```
Sorted by: id
```

```
. set matsize 800
```

Current memory allocation

settable	current value	description	memory usage (1M = 1024k)
set maxvar	5000	max. variables allowed	1.909M
set memory	50M	max. data space	50.000M
set matsize	800	max. RHS vars in models	4.950M

			56.859M

```
. svyset [weight=weights], strata(region) psu(area)
```

(sampling weights assumed)

```
pweight: weights
VCE: linearized
Single unit: missing
Strata 1: region
SU 1: area
FPC 1: <zero>
```

. svy: mean agecar

(running mean on estimation sample)

Survey: Mean estimation

Number of strata = 5 Number of obs = 1161
 Number of PSUs = 54 Population size = 331186
 Design df = 49

```
-----+-----
```

	Mean	Linearized Std. Err.	[95% Conf. Interval]
agecar	5.802543	.	.

```
-----+-----
```

Note: missing standard error because of stratum with single sampling unit.

. tab1 carhuman unworthy alcohol

-> tabulation of carhuman

carhuman	Freq.	Percent	Cum.
0	771	66.41	66.41
1	390	33.59	100.00
Total	1,161	100.00	

-> tabulation of unworthy

unworthy	Freq.	Percent	Cum.
0	483	41.60	41.60
1	678	58.40	100.00
Total	1,161	100.00	

-> tabulation of alcohol

alcohol	Freq.	Percent	Cum.
0	252	21.71	21.71
1	909	78.29	100.00
Total	1,161	100.00	

. svy: tab carhuman alcohol, obs

(running tabulate on estimation sample)

Number of strata = 5 Number of obs = 1161
 Number of PSUs = 54 Population size = 331186.46
 Design df = 49

```
-----+-----
```

carhuman	alcohol		Total
	0	1	
0	.1503	.5758	.7261
	174	597	771
1	.0412	.2327	.2739
	78	312	390
Total	.1915	.8085	1

```
-----+-----
```

```
| 252 909 1161
```

```
-----  
Key: cell proportions  
      number of observations
```

```
Pearson:  
Uncorrected chi2(1) = 4.8037  
Design-based F(., .) = . P = .
```

Note: missing test statistics because of stratum with single sampling unit.

. svy: logit carhuman curve brakes unworthy alcohol, or

(running logit on estimation sample)

Survey: Logistic regression

```
Number of strata = 5          Number of obs = 1161  
Number of PSUs = 54         Population size = 331186.46  
Design df = 49  
F( 0, 49) = .  
Prob > F = .
```

```
-----  
carhuman | Odds Ratio   Linearized  
          |              Std. Err.   t    P>|t|   [95% Conf. Interval]  
-----+-----  
curve | 1.351633      .          .    .    .    .  
brakes | 2.444719      .          .    .    .    .  
unworthy | 1.340772      .          .    .    .    .  
alcohol | .9392623      .          .    .    .    .  
-----
```

Note: missing standard errors because of stratum with single sampling unit.

Basic STATA commands for the analysis of complex survey data sets

```
. use c:\students\sello3.dta, clear  
  
. set matsize 800  
  
. svyset [weight=weights], strata(strata) psu(psu)  
  
. tab survival (Frequency table)  
  
. svy: mean agemoth
```

```
. tab1 survival lit childU5 sex famsize toiletrec  
delivery
```

Pearson's chisquare test of association between
mother's literacy status and survival

```
. svy: tab lit survival, obs
```

Binary logistic regression analysis for complex
survey data sets

```
. svy: logit survival lit childU5 sex famsize  
toiletrec delivery contuse ancvisit dbfcat, or
```

How to obtain adjusted odds ratios

```
. svy: logit survival lit childU5 sex famsize  
toiletrec delivery contuse ancvisit dbfcat, or  
. adjust sex famsize, by(contuse ancvisit) exp ci
```

[95% Confidence Interval]

How to obtain design effects for assessing the
adequacy of the fitted logistic regression model

```
. estat effects
```