METHODOLOGY STUDY FOR DETERMINING THE OPTIMUM RECALL PERIOD FOR THE REPORTING OF MOTOR VEHICLE ACCIDENTAL INJURIES

William S. Cash and Abigail J. Moss
National Center for Health Statistics

I. Introduction

Since the beginning of the National Health Interview Survey in 1957, annual estimates on the incidence of injuries resulting from all types of accidents have been obtained. The most recent injury data available are for the period July 1966-June 1967; the data indicate that during this period, an estimated 51.8 million persons, 26.9 per 100 persons in the civilian, noninstitutional population, were injured. Of this number, 3.5 million or 1.8 per 100 persons were injured in moving motor vehicle accidents. Even though the number of persons injured in moving motor vehicle accidents constitutes only 6.8 percent of the total injured population, these data have been of particular interest to data consumers. One reason for this interest is that motor vehicle injures are often of a more serious nature than other types of injures. For example, the proportion of motor vehicle injuries resulting in activity restriction and bed disability is markedly higher than for other types of injuries.[1]

During the past two years, the National Center for Health Statistics has experienced an increased demand for more reliable and detailed statistics on motor vehicle injuries and other factors relating to motor vehicle accidents. Part of this demand was the result of a greater public awareness of the paucity of motor vehicle safety standards employed at that time and the high number of traffic fatalities and personal injuries resulting from motor vehicle accidents. Since the Center had not previously collected motor vehicle injury data in sufficient detail to satisfy the requests being made, a decision was made to obtain more detailed information on motor vehicle injuries on the 1968 Health Interview Survey questionnaire. Prior to the actual data collection phase however, it appeared that an evaluation study should be conducted for the purpose of establishing new estimating procedures for motor vehicle data. This was thought to be necessary since the collection and sampling procedures used earlier to estimate the annual incidence of injuries would result in an exceedingly high sampling error, if used to derive annual estimates for more detailed motor vehicle data. This report describes the methodological aspects of this special study which took place between February and May 1967 and presents the findings which were later incorporated by the Health Interview Survey.

In the past, estimates on the incidence of all types of injuries have been obtained by collecting data among sample persons on only those injuries that occurred during the two-week period preceding the interview and then inflating the frequencies to obtain annual estimates. The collection of injury data has been limited to a two-week recall period primarily because some specific kinds of injuries a person may receive have such little impact that respondents may forget to report them if much time has elapsed between the date of the accident and the interview. The degree of impact an injury has on an individual could be expected to vary depending upon: (1) the severity of the injury involved, and (2) the circumstances of the accident which caused the injury or, more specifically, the type of accident that occurred. Therefore, if one were to hypothesize that injury-producing motor vehicle accidents have a greater impact on an individual than injuries obtained from some other kinds of accidents, it could be assumed that a respondent would be able to remember this type of injury for a longer time period and report it in a household interview even when a recall period longer than two weeks was used. By increasing the length of the recall period for motor vehicle injuries, the number of injuries reported would be increased. This, in turn, would have the effect of decreasing the sampling error, making it feasible to collect and publish motor vehicle injury data in greater detail.

With these considerations in mind, this evaluation study was specifically designed to answer the following two questions: (1) can the recall period for injuries resulting from motor vehicle accidents be increased without greatly affecting the respondent's ability to report such occurrences, and if this is possible, (2) what is the optimum length of recall for the reporting of motor vehicle injuries?

II. The Study Design

Description of the Survey Procedure. After considering the various alternatives, it was determined that the best method for evaluating the optimum recall period for the reporting of motor vehicle injuries was a record check study. Briefly, the Motor Vehicle Evaluation Study would consist of interviewing a sample of persons known to have been in an injury-producing motor vehicle accident at some time during the twelve-month period preceding the interview. Accident information obtained from the respondent at the time of the interview would then be compared with comparable data recorded on an official accident report form. Final analysis of the data would consist of a comparison between a person's injury status as recorded on the accident record and on the questionnaire used in the interview, as well as a comparison of the reporting of other details relating to the accident. Of primary interest for evaluation however, would be the relationship between the ability of the respondent to report motor vehicle injuries and the length of time between the motor vehicle accident and the interview.

The Sample. The Motor Vehicle Evaluation Study was conducted in the Research Triangle Area of North Carolina where the Division of Health Interview Statistics has an experimental field interview unit established for the purpose of conducting methodological studies of this kind. The sample design used for this study had the following features. First, the North Carolina Department of Motor Vehicles provided the Health Interview Survey staff with accident record punch cards for those accidents occurring in Durham, Orange, and Wake counties in that State during February 1966 to February 1967 which met all of the specifications listed below:

1. one or more of the persons involved in the accident were residents of Durham, Orange, or Wake counties

2. one or more persons in the accident were injured

3. one or more persons survived the accident

These punch cards were then divided into three strata according to the time interval between the date of the accident and the interview: Stratum I, less than 3 months; Stratum II, 3-6 months; and Stratum III, 7-12 months. Within these strata, the cards were sorted by:

1. whether or not a legal violation was involved, and

2. the most severe injury sustained in the accident as reported by the police officer who completed the official accident report form.

The sample for the Motor Vehicle Evaluation Study was then drawn by using a simple random (systematic) sample by strata. The sampling fraction for Strata I and III was 1/6 and for Stratum II was 1/5. In order to detect differences at the .05 significance level, it was estimated that approximately 500 accidents would need to be selected for this study. The actual number of households finally interviewed, however, was considerably more than that since a certain proportion of accidents involved two or more persons residing at different addresses.

The Motor Vehicle Accident Report Form. After the sample was drawn, the North Carolina Department of Motor Vehicles was requested to provide the Division of Health Interview Statistics with a copy of the original accident report form for each accident falling in the sample. The report forms contained the name and address of the driver(s) involved regardless of whether he was injured and the name and address of all other persons in the accident who were either injured or killed. The record also contained a classification of the type of injury each injured person sustained and specific details of the accident. The three injury classifications were as follows:

1. A type A injury was described as a visible sign of injury, such as a bleeding wound or distorted limb, or where the person had to be carried from the accident scene;

2. A type B injury included other visible injuries or bruises, abrasions, swelling, limping and so forth;

3. A type C injury involved no visible sign of injury but the person experienced momentary unconsciousness or complained of pain.

The Field Operation. Once the motor vehicle accident report forms were received, the addresses of all persons residing in the three-county area were abstracted from the record. These sample addresses were then grouped into interview segments according to geographical proximity to one another. Prior to the actual interview, an advance letter was sent to each sample address informing the residents that they would be contacted by an interviewer from the U. S. Public Health Service who would ask them questions about the health of their family and other health-related items.

Interviewing was carried out by nine interviewers from the Health Interview Survey field staff over an eleven-week period, from February 20 through May 5, 1967. The interviewers received their training from staff members of the Division of Health Interview Statistics. The questionnaire the interviewers used for this study was a substantially shortened version of the schedule implemented in the ongoing program of the Health Interview Survey in 1968. In addition to the basic questionnaire, a motor vehicle accident supplement was completed for each reported accident; the supplement contained the detailed questions about the types of injuries sustained and other particulars about the accident.

As can be seen from Table A, an attempt was made to interview 939 sample households; of this number, 809 households were finally interviewed. Since the sample for this study was selected from records representing accidents occurring as much as a year preceding the interview, it was expected that some proportion of sample persons would be lost because they had moved from the original address. To minimize this kind of loss, however, a follow-up procedure was initiated:

> At the close of each interview, several questions were asked to obtain the name, and, if possible, the present address of any person who had resided in that household at any time during the past 12 months but who was not living there at the time of the interview. These questions had to be asked in each household since the interviewers, in most instances, were not given the name of the sample person, and, consequently, did not know when they were interviewing a sample family. An attempt

was eventually made to interview those sample persons reported to have moved from the original address provided: (1) they still lived within the three-county area, and (2) the new address obtained at the original household contained sufficient information to locate the person.

Table A. Number and percent distribution of completed interviews of sample and non-sample households by place of interview and number and percent distribution of non-interviews.

Interview Status	Number	Percent Distributions		
Total households	939	100.0		
Completed interviews	809	86.2	100.0	
Households yielding sample person:	640	68.2	79.1	100.0
at original address	625	66.6	77.3	97.7
at follow-up address	15	1.6	1.8	2.3
Households not yielding sample person:	169	18.0	20.9	100.0
Interview completed at original address	165	17.6	20.4	97.6
Interview completed at follow-up address	4	.4	.5	2.4
Non-interviews	130	13.8		

Of the 640 completed interviews (table A) resulting in a sample person, 15, or 2.3 percent, were the result of interviews conducted at a follow-up address. This percentage seems small when compared with the 625, or 97.7 percent, of the households yielding a sample person at the original address. However, the additional effort that went into locating these few sample persons seems worthwhile when considering that of the 19 interviews conducted at a follow-up address, slightly over three-fourths of them yielded a sample person.

About 14 percent of the households were never interviewed. This compares favorably with the percentage of type A, B, and C non-interviews in the ongoing Health Interview Survey. Table B shows a breakdown of these non-interviewed households by reason for non-interview.

The largest single factor contributing to the over-all non-interview rate was that in 23 households it was learned prior to the interview that the sample person no longer lived at the sample address. Another 17 households were never located by the interviewers. The problem of locating households occurred because, in some instances, the accident record contained an inaccurate or incomplete address. In this study, interviewers had the most difficulty finding households with rural addresses. The local Post Offices, however, provided some help in locating these households, minimizing the number of addresses that could not be located.

For the most part, the two types of non-interviews described above do not occur with any great frequency in the Health Interview Survey, and may explain why the non-interview rate in this study was higher than that found in the ongoing Survey. The percentage distributions for the other types of non-interviews, such as refusals, temporary absences, and demolished residences, were quite similar in the special study and the Survey.

Table B. Number and percent distribution of completed interviews and non-interviews by reason for non-interview.

Interview Status	Number	Percent Distribution	
Total households	939	100.0	
Completed interviews	809	86.2	
Non-interviews	130	13.8	100.0
Type A	51	5.4	39.2
Refusal	18	1.9	13.8
Not at home	22	2.3	16.9
Other	11	1.2	8.5
Type B	27	2.9	20.8
Vacant	18	1.9	13.8
Other	9	1.0	6.9
Type C	50	5.3	38.5
Not sample household	23	2.4	17.7
Could not locate house	17	1.8	13.1
Other	10	1.1	7.7
Non-interview status unknown	2	.2	1.5

III. Analysis of Data

The analysis of the record case study is based on the data in tables 1-8, and is discussed below. The solution to the problem of determining the optimum recall period will be treated in a later section of this report.

Table 1 shows that a total of 590 sample persons were interviewed. A sample person is defined as any person listed on the motor vehicle record who resided within the three-county area at the time of the accident. This includes all drivers, whether injured or not as indicated on the record, and all injured passengers. Also, any person reporting an injury, regardless of his injury status on the accident report form, is defined as a sample person.

Other facts of interest which are apparent from the data in table 1 are:

1. Eighty-two sample persons, or 13.9 percent of the total sample persons interviewed, did not report the accident.

2. The non-reporting of accidents increases as the time between the date of accident and interview increases. The non-reporting ranges from 3.4 percent for less than three months to a maximum of 27.3 percent for the interval of nine-twelve months. The obvious reason for this trend results from an increased inability to recall the occurrence of a motor vehicle accident as the time between the date of accident and the date of interview increases.

Of the 590 sample persons interviewed, the motor vehicle record indicates that 377 persons, or 63.9 percent, were injured (table 2). There are several points of interest evident from this table:

1. For the recall period of less than three months, 87.3 percent of the 377 injured persons interviewed reported the injury sustained in the accident. This compares with 78.8 percent for a recall period of less than six months, and 75.1 percent for less than twelve months.

2. Fifty-one sample persons, or 13.5 percent of the sample persons reported as injured on the motor vehicle record, reported the occurrence of the accident but did not report the injury. It is possible that the injury classification on the record may not be absolutely correct, and that not all of the 51 sample persons who reported the accident but failed to report the injury actually sustained an injury in the accident. However, since the record is being used as a criterion to estimate the ability of a respondent to report motor vehicle injuries, the bias introduced by inaccuracies in the record must be accepted. If the assumption is not made that the record is correct, no valid foundation exists on which the determination of the optimum recall period can be made.

3. Forty-three persons, or 11.4 percent of the sample persons classified on the record as being injured, did not report the sample accident. This percentage increased as the recall period became longer.

It can be seen from table 3 that, of the 590 sample persons who were involved in accidents and were interviewed, 213, or 36.1 percent, were reported as not injured on the motor vehicle record:

1. Thirty-four sample persons, or 16.0 percent of the 213 persons in this group, reported an injury when, in fact, the record indicated no injury. Most of these injuries were reported within a six-month recall period. This reporting trend may indicate that these injuries were minor, and were less likely to be reported as the recall period was extended beyond six months.

2. Thirty-nine persons, or 18.8 percent of the noninjured sample persons, did not report the accident. When this percentage is compared with the 11.4 percent of the injured persons who did not report the accident (shown in table 2), it seems that a respondent is more likely to report an accident if he received an injury in the accident.

The reporting of the accident and injury in the interview, by type of injury received (as indicated on the motor vehicle record), is shown in table 4. The following statistics are of importance:

1. For the recall period of less than twelve months, 85.5 percent of the type A injuries were reported, as compared to 67.8 percent type B, and 67.2 percent type C injuries (for definitions of type A, B, and C injuries, see page 2). This difference is significant, and can be interpreted as a result of the degree of severity of injury which is inherent in the definition of type A, B, and C injuries. This trend is also apparent for recall periods of three months and six months.

2. The reporting in the interview of type B and C injuries appears equally good. This similarity was unexpected, since type B injuries are, by definition, more severe than type C injuries.

3. For the recall period of less than twelve months, 10.7 percent of the sample persons who incurred type A injuries did not report the accident, as compared to 13.3 percent for type B, and 10.9 percent for type C injuries. These percentages indicate that the reporting of the accident is independent of the type of injury received. However, as indicated in point one above, the reporting of the injury itself is dependent on the type of injury received.

In this study, sample persons were classified into two response groups according to the following criteria. A sample person was classified as a self respondent if he or some other person(s) involved in the accident participated in the interview. If this condition is not met, the sample person is considered as having a proxy respondent.

The reporting and non-reporting of the accident and injury are shown by respondent status in table 5. Of particular interest in table 5 is the 3.9 percent of all self respondents, compared to 11.7 percent of the proxy respondents, who did not report the accident when the sample person was injured. Also, since self respondents are generally able to report most events more accurately than proxy respondents, it was surprising that 6.9 percent of self respondents, compared to only 4.3 percent proxy respondents, reported an injury

for the sample person when the accident record indicated none. This reporting pattern probably occurred because a few sample persons who were not classified on the accident record as being injured actually received minor injuries, rather than because proxy respondents reported this item more accurately than did self respondents. Reporting differences for self and proxy respondents for the other categories in table 5 are small.

Inaccuracies in reporting the date of accident (by time interval) among sample persons reporting the accident are shown in table 6. The following points are of interest:

1. From table 1 it can be seen that, according to the record, 119 sample persons had an accident which occurred within a three-month period prior to interview. Of this number, 115 persons, or 96.6 percent, reported the accident. Of these 115 sample persons, 6 persons, or 5.2 percent, reported the accident as occurring in the interval three-six months prior to interview. This underreporting for the interval less than three months is counterbalanced by the reporting of 16 sample persons, or 8.5 percent of the sample persons, who had an accident in the interval three-six months prior to interview, but who reported in the interview that the accident occurred less than three months prior to the time of interview.

2. For the recall period less than six months, 5 persons, or 1.7 percent of the sample persons reporting the accident, reported it as occurring in the interval six-nine months prior to interview. This compares with 18 sample persons, or 8.7 percent, of the sample persons who reported the accident as occurring in the interval less than six months, when, according to the record, the accident occurred six-twelve months prior to interview.

3. Due to delays in interviewing, 59 sample persons were interviewed more than twelve months after the date of the accident. Of this number, 5 sample persons, or 8.5 percent of the 59 sample persons, reported the accident as occurring within the past twelve months.

4. The over-all pattern indicates that a certain proportion of the people who reported the sample accident, reported it as occurring earlier than the actual date of the accident. This phenomenon occurs at a slightly higher rate than the proportion of people who reported the occurrence of the accident on a date later than when it actually occurred. The net difference appears insignificant when examined for the three recall periods of less than three months, less than six months, and less than twelve months. For this reason, analysis of the optimum recall period will not be based on the bias in reporting of the date of the accident as shown in this table.

Information on the completeness of reporting of the accidents in all sample households is shown in tables 7 and 8, by interval since the occurrence of the accident and by respondent status:

1. Table 7 shows that 532 sample households resulted in an interview which yielded a sample person. The 532 households yielded 590 sample persons (table 1).

2. In 79 sample households, or 14.8 percent of the 532 households, the accident was not reported. This proportion compares with 13.9 percent of all sample persons who did not report the accident (table 3). Nonreporting of the accident increases as the recall period increases. The percentage ranges from 2.8 percent for recall of less than three months to 30.7 percent for the interval of 9-12 months.

3. Of the 532 sampled households, 287, or 53.9 percent, were self-responding households (table 8). A household is defined as self-responding if at least one person who responded in the interview was also in the accident, whether or not he was injured. If a household could not be classified in the self-responding category, then the household was defined in the proxy category. Of the 287 self-responding households, 29, or 10.1 percent, did not report the accident. This compares with 20.4 percent in the proxy-responding households. This difference is significant, and indicates the magnitude of bias which might result when information is obtained from a proxy respondent versus a self respondent.

IV. Determining the Optimum Recall Period

The National Center for Health Statistics has collected motor vehicle injury data in its National Health Interview Survey for the year 1968. The question asked of each respondent was: During the past twelve months, have you been in a motor vehicle accident, either as a driver, passenger, or a pedestrian? The data from this question have been processed and are available for analysis. National estimates of persons injured in moving motor vehicle accidents, as well as information about factors relating to the accident, are to be published. The purpose of the record case study is to assist in the determination of the recall period to be used in the interview survey that will give the most reliable estimate of P_{12}, the true proportion of motor vehicle injuries which have occurred in the United States during 1968. The concept and definition of a recall period have been discussed in the Analysis of Data section. The procedure for estimating the proportion of motor vehicle injuries which have occurred in the United States during 1968 is directly related to the recall period selected.

An example will best illustrate this relationship. If a less than three-month recall period is selected for estimating the total number of

motor vehicle injuries occurring within the year, the procedure would be to estimate the total number of injuries occurring within a three-month interval and inflate this estimate to represent the total number of motor vehicle injuries occurring within the year. A reported injury in the Health Interview Survey is within a three-month interval if the respondent reported the injury as occurring within the three months prior to the date of interview. If the respondent reported the injury as occurring more than three months prior to date of interview, this injury would not be inflated in the estimation of the total number of motor vehicle injuries. A similar definition would hold for any other recall period period. The recall periods which will be considered in this analysis are: less than three months, less than six months, and less than twelve months.

For any recall period, there are two components of precision which must be carefully examined. The first of these is the variance of the estimator $\hat{P}_{12}$, and the second is the biasness of $\hat{P}_{12}$. $\hat{P}_{12}$ is the estimated proportion, or rate, of motor vehicle injuries occurring in the United States during the year 1968. There are two properties of variance and bias which are of importance when considering the three recall periods:

1. The variance of $\hat{P}_{12}$ decreases when the longer recall period is used. That is, a 12-month recall period results in a smaller variance for $\hat{P}_{12}$ when compared with the variance which results from using a six-month, or a three-month recall period.

2. The bias of $\hat{P}_{12}$ increases when the longer recall period is used. Bias is measured by the proportion of people who fail to report a motor vehicle injury. Bias increases since the ability of a respondent to recall a motor vehicle injury decreases as the recall period is lengthened.

The technique for determining the optimum recall period consists of the selection of the recall period which minimizes the sum of the variance component of $\hat{P}_{12}$ and the square of the bias of $\hat{P}_{12}$. In statistical terms, this method of optimization is referred to as the minimum mean-squared error, MSE.

Estimates of Variance, Mean-squared Error and Relative Root Mean-square Error. The sample size, estimated probability of injury, variance based on assumption of independence and non-independence, and the estimated bias of $\hat{P}_{12}$ squared are shown in table 9. The subscripts 3, 6, and 12 refer to three-month, six-month, and twelve-month recall periods, respectively. The following comments and explanations are needed:

1. The sample size N = 134,000. The sample size represents the estimated total number of people interviewed in the National Health Interview Survey in 1968, and remains constant for each of the three recall periods.

2. The true probability of a person receiving a motor vehicle injury for the entire year 1968 is denoted by P_{12}. It is assumed that this probability is uniform over the twelve-month period, and hence: $P_3 = \frac{1}{4}P_{12}$, and $P_6 = \frac{1}{2}P_{12}$, where P_3 and P_6 denote the probability of a person receiving a motor vehicle injury in a three-month and six-month time interval, respectively.

3. The variance of the estimated probability of injury in the past twelve months is shown for each of the recall periods. The variance which will be used in the analysis is the variance due to lack of independence. Independence is not satisfied in the National Health Survey, since the basic sampling unit is a household. That is, all of the respondents in a household would tend either to report or not report the accident and injuries. The variance of $\hat{P}_{12}$ due to lack of independence is expected to be at least twice as large as the variance of $\hat{P}_{12}$ if independence could be assumed.

4. The estimated bias of $\hat{P}_{12}$ squared is shown for each recall period. The factors K_3, K_6, and K_{12} are estimates of the bias in reporting motor vehicle injuries for three-month, six-month, and twelve-month recall periods, respectively. Estimates of the bias components from table 2 are: $K_3 = .127$, $K_6 = .212$, and $K_{12} = .249$. These estimates are based on the proportion of people who were reported as being injured on the motor vehicle record but failed to report the injury when interviewed.

5. The MSE of $\hat{P}_{12}$ by definition is equal to the variance of $\hat{P}_{12}$ plus the square of the bias of $\hat{P}_{12}$.[2] From table 9, the MSE of $\hat{P}_{12}$ can be determined for each recall period. Variance due to lack of independence is used in the MSE formula.

6. The RRMSE of $\hat{P}_{12}$ can be determined for each recall period from the formula:

$$\text{RRMSE}(\hat{P}_{12}) = \frac{\sqrt{\text{MSE}(\hat{P}_{12})}}{P_{12}} \times 100\%$$

It should be realized that the recall period which results in the minimum MSE of $\hat{P}_{12}$ will also result in the minimum RRMSE of $\hat{P}_{12}$. This can be seen be examining the RRMSE formula. The RRMSE shows the error of the estimate $\hat{P}_{12}$ as a percentage of the true proportion, P_{12}. In addition to selecting the recall period which gives the minimum RRMSE of $\hat{P}_{12}$, a further requirement is that the RRMSE of $\hat{P}_{12}$ for this recall period shall not exceed 25 percent. A RRMSE of 25 percent or less is considered an acceptable level for showing estimates of proportions or totals.

7. The variance component, X_1, of the RRMSE of $\hat{P}_{12}$ is the relative standard error of $\hat{P}_{12}$, or

$$X_1 = \frac{\sqrt{\text{Var}\ (\hat{P}_{12})}}{P_{12}} \times 100\%$$

It should be realized that the variance component is identical to the RRMSE of $\hat{P}_{12}$ if $\hat{P}_{12}$ is an unbiased estimator of P_{12}.

8. The bias component, X_2, of the RRMSE of $\hat{P}_{12}$ is the difference between the RRMSE of $\hat{P}_{12}$ and the variance component, or

$$X_2 = \text{RRMSE}(\hat{P}_{12}) - \frac{\sqrt{\text{Var}\ (\hat{P}_{12})}}{P_{12}} \times 100\%$$

In order to determine which recall period results in a minimum RRMSE from the RRMSE equation, it can be seen that we must assume a value for P_{12}. Data collected by the National Center for Health Statistics for the period July 1966-June 1967 show that an estimated 3.5 million persons, or a rate of 1.8 persons per 100, were injured in moving motor vehicle accidents. An estimate of P_{12} based on these data is 1.8×10^{-2}. Since it is desirable to show not only the estimated proportion of people injured, but also a categorization of this proportion by such characteristics as age, sex, driver status, residence, region, severity of accident, and possibly other variables, it is necessary to take these into consideration in the methodology, because an estimate of P_{12} based on these characteristics would be much smaller than 1.8×10^{-2}. For this reason, the optimum recall period is shown as a function of P_{12}. The following inequalities are solved for P_{12}.

1.1 $\quad \text{RRMSE}\ (4\ \hat{P}_3) \leq \text{RRMSE}\ (2\ \hat{P}_6)$

1.2 $\quad \text{RRMSE}\ (4\ \hat{P}_3) \leq \text{RRMSE}\ (2\ \hat{P}_{12})$

1.3 $\quad \text{RRMSE}\ (2\ \hat{P}_6) \leq \text{RRMSE}\ (\hat{P}_{12})$

Let the solution of equation 1.1 for P_{12} be P'_{12}. This implies that a three-month recall period results in a smaller RRMSE, when compared to a six-month recall for all values of P_{12} less than or equal to P'_{12}. A similar interpretation holds in equation 1.2, comparison of three-month recall to a twelve-month recall; and equation 1.3, which compares a six-month recall to a twelve-month recall. By simultaneously considering equations 1.1, 1.2, and 1.3, and their solutions, a graph can be constructed showing the values of P_{12} which result in a minimum RRMSE for each of the recall periods. However, solutions in terms of P_{12} have very little intuitive meaning. For this reason solutions are shown in terms of T_{12}, where T_{12} is the population size of injured persons which results when P_{12} is inflated to represent the total United States population, that is, $T_{12} = 200 \times 10^6 \times P_{12}$.

Graph 1. The recall period resulting in the minimum RRMSE for specific population sizes of injured persons.

Twelve-month Recall Period	Six-month Recall Period	Three-month Recall Period	
174,000	195,000	207,000	T_{12}

Graph 1 above shows the population sizes of injured persons which result in a minimum RRMSE for each of the recall periods. The following statistics are of interest:

1. Graph 1 shows that a twelve-month recall period results in a minimum RRMSE for estimates on specific injured populations of size less than 174,000. For estimates ranging from size 174,000 to 207,000, a six-month recall period results in a smaller RRMSE when compared to a twelve-month recall period of injured populations of size greater than 195,000. For estimates larger than 207,000, a three-month recall period yields the minimum RRMSE over both the six-month and twelve-month recall periods.

2. A twelve-month recall period yields the minimum RRMSE for estimates on small population sizes. As the population size increases, the six-month recall period becomes optimum over the twelve-month recall period. This occurs at a population of size 154,000. Eventually, the population size increases to a point (207,000) where the three-month recall yields the minimum RRMSE.

However, graph 1, above, does not show the actual value of the RRMSE, but only the population sizes for which each recall period yields the minimum RRMSE. Table 10 shows the value of the RRMSE, the variance component, and the bias component for each recall period as the population varies in size from 25 thousand to 5 million. Based on data from this table and graph 1, the optimum recall period for estimating the total number of persons injured in motor vehicle accidents is the recall period of less than three months. The following reasons support this decision:

1. The RRMSE of estimates larger than 207,000 is a minimum for the less than three-month recall period. As the size of the estimates increases, table 10 shows that the RRMSE based on a recall period of 12 months decreases slightly from 27.7 percent to 25.0 percent. For a less than six-month recall period, this decrease is from 27.3 percent to 21.5 percent. The largest decrease occurs in the three-month recall period, where the RRMSE declines from a level of 27.5 percent to 13.6 percent.

2. Estimates of greatest interest are for populations of size greater than 207,000. Indeed, the single most important estimate is the

total number of moving motor vehicle injuries, which is estimated to be nearly 4 million. The RRMSE for an estimate of 4 million is 25.0, 21.5, and 13.8 percent for a twelve-month, six-month, and three-month recall period, respectively. The contrast in these three percentages led to the selection of a recall period covering the three-month interval preceding the week of interview.

3. Data collected in the Health Interview Survey for the period July to December 1967 have been processed and estimates of the total number of persons injured in motor vehicle accidents within the year have been made using each of the recall periods. The estimated total number of persons injured is 3.2, 2.7, and 2.4 million based on three-month, six-month, and twelve-month recall periods, respectively. From a comparison of these estimates, it seems that the bias component of the RRMSE, which is a function of the ability of a respondent to recall a motor vehicle injury, increases over time at a rate greater than estimated from this methodology study. Hence, it appears that the results of this study, which led to the selection of a three-month recall period, are conservative.

4. The variance component and bias component have certain effects on the value of the RRMSE. As the estimated number of persons injured increases, the variance component of the RRMSE decreases, the bias component increases, and the RRMSE decreases (figure 2).

REFERENCES

[1] National Center for Health Statistics, Current Estimates from the Health Interview Survey, PHS Publication No. 1000, Series 10, No. 43, p. 3, Washington, D.C., January 1968.

[2] Graybill, F. A., and Mood, A. M. (1963), Introduction to the Theory of Statistics, New York, McGraw-Hill.

Table 1. Reporting and non-reporting of sample accident for all sample persons interviewed.

Time Lapse Between Date of Accident and Date of Interview	All Sample Persons					
	Total		Reported Sample Accident		Did not report Sample Accident	
	Frequency	Percent Dist.	Frequency	Ratio Dist.	Frequency	Ratio Dist.
Total	590	100.0	508	86.1	82	13.9
Less than 3 months	119	20.2	115	96.6	4	3.4
3 - 6 months	209	35.4	187	89.5	22	10.5
6 - 9 months	119	20.2	102	85.7	17	14.3
9 - 12 months	143	24.2	104	72.7	39	27.3
Less than 6 months	328	55.6	302	92.1	26	7.9
6 - 12 months	262	44.4	206	78.6	56	21.4
Less than 12 months	590	100.0	508	86.1	82	13.9

Table 2. Reporting and non-reporting of sample accident and injury for all sample persons who were injured on Motor Vehicle Record.

Time Lapse Between Date of Accident and Date of Interview	Sample Person Injured on Motor Vehicle Record									
	Total		Sample Accident Reported						Sample Accident Not Reported	
					Reported Injury		Did not Report Injury			
	Frequency	Percent Dist.	Frequency	Ratio Dist.	Frequency	Ratio Dist.	Frequency	Ratio Dist.	Frequency	Ratio Dist.
Total	377	100.0	334	88.6	283	75.1	51	13.5	43	11.4
Less than 3 months	71	18.8	70	98.6	62	87.3	8	11.3	1	1.4
3-6 months	141	37.4	127	90.1	105	74.5	22	15.6	14	9.9
6-9 months	71	18.8	64	90.1	57	80.3	7	9.9	7	9.9
9-12 months	94	24.9	73	77.7	59	62.8	14	14.9	21	22.3
Less than 6 months	212	56.2	197	92.9	167	78.8	30	14.2	15	7.1
6-12 months	165	43.8	137	83.0	116	70.3	21	12.7	28	17.0
Less than 12 months	377	100.0	334	88.6	283	75.1	51	13.5	43	11.4

Table 3. Reporting and non-reporting of sample accident and injury for all sample persons <u>not</u> injured on Motor Vehicle Record.

Time Lapse Between Date of Accident and Date of Interview	Sample Person Not Injured on Motor Vehicle Record									
	Total		Sample Accident Reported						Sample Accident Not Reported	
					Did Not Report Injury		Did Report Injury			
	Frequency	Percent Dist.	Frequency	Ratio Dist.	Frequency	Ratio Dist.	Frequency	Ratio Dist.	Frequency	Ratio Dist
Total	213	100.0	174	81.2	140	65.7	34	16.0	39	18.8
Less than 3 months	48	22.5	45	93.8	30	62.5	15	31.3	3	6.2
3-6 months	68	31.9	60	88.2	50	73.5	10	14.7	8	11.8
6-9 months	48	22.5	38	79.2	34	70.8	4	8.3	10	21.8
9-12 months	49	23.0	31	63.3	26	53.1	5	10.2	18	36.7
Less than 6 months	116	54.5	105	90.5	80	69.0	25	21.6	11	9.5
6-12 months	97	45.5	69	71.1	60	61.9	9	9.3	28	28.9
Less than 12 months	213	100.0	174	81.2	140	65.7	34	16.0	39	18.8

Table 4. Reporting of accident and injury by type of injury on Motor Vehicle Record.

Time Lapse Between Date of Accident and Date of Interview	Injury Classification on Motor Vehicle Record											
	Type A				Type B				Type C			
	Total	Reported Injury	Reported Accident Only	Did Not Report Accident	Total	Reported Injury	Reported Accident Only	Did Not Report Accident	Total	Reported Injury	Reported Accident Only	Did Not Report Accident
Less than 3 months												
Number	29	27	2	-	13	11	2	-	29	24	4	1
Percent Dist.	100.0	93.1	6.9	-	100.0	84.6	15.4	-	100.0	82.6	13.7	3.7
3-6 months												
Number	59	53	1	5	34	23	7	4	48	29	14	5
Percent Dist.	100.0	89.8	1.7	8.5	100.0	67.6	20.3	11.8	100.0	60.4	29.2	10.4
6-9 months												
Number	27	24	1	2	15	13	1	2	28	20	5	3
Percent Dist.	100.0	88.9	3.7	8.4	100.0	81.3	6.2	12.5	100.0	71.4	17.9	10.7
9-12 months												
Number	44	32	2	10	27	14	7	6	23	13	5	5
Percent Dist.	100.0	72.7	4.5	22.8	100.0	51.8	26.0	22.2	100.0	56.6	21.7	21.7
Less than 6 months												
Number	88	80	3	5	47	34	9	4	77	53	18	6
Percent Dist.	100.0	90.9	3.4	5.7	100.0	72.3	19.1	8.5	100.0	68.8	23.8	7.4
6-12 months												
Number	71	56	3	12	43	27	8	8	51	33	10	8
Percent Dist.	100.0	78.9	4.2	16.9	100.0	62.8	18.6	18.6	100.0	64.7	19.6	15.7
Less than 12 months												
Number	159	136	6	17	90	61	17	12	128	86	28	14
Percent Dist.	100.0	85.5	3.8	10.7	100.0	67.8	18.9	13.3	100.0	67.2	21.9	10.9

Table 5. Number and percent distribution of reporting and non-reporting of accident and injury by respondent status.

Respondent Status	Total	Sample Person Did Not Report Accident		Sample Person Reported Accident Only		Sample Person Reported Injury	
		Injured on Record	Not Injured on Record	Injured on Record	Not Injured on Record	Injured on Record	Not Injured on Record
Total Persons	590	43	39	51	140	283	34
Self Respondents							
Number	333	13	18	29	80	170	23
Percent Dist.	100.0	3.9	5.4	8.7	24.0	51.1	6.9
Proxy Respondents							
Number	257	30	21	22	60	113	11
Percent Dist.	100.0	11.7	8.2	8.5	23.3	44.0	4.3

Table 6. Error in reporting date of accident by time interval for all sample persons reporting sample accident.

Time Lapse Between Date of Accident and Date of Interview	Total	Sample Accident Reported as Occurring in the Following Time Interval from Date of Interview				
		Less than 3 months	3-6 months	6-9 months	9-12 months	DK Date Given
Total*	508	125	189	100	90	4
Less than 3 months	115	109	6	-	-	-
Percent distribution	100.0	94.8	5.2	-	-	-
3 - 6 months	187	16	165	5	-	1
Percent distribution	100.0	8.5	88.2	2.7	-	.6
6 - 9 months	102	-	14	83	3	2
Percent distribution	100.0	-	13.7	81.4	2.9	2.0
9 - 12 months	104	-	4	12	87	1
Percent distribution	100.0	-	3.8	11.5	83.7	1.0
Less than 6 months	302	125	171	5	-	1
Percent distribution	100.0	41.4	56.6	1.7	-	.3
6 - 12 months	206	-	18	95	90	3
Percent distribution	100.0	-	8.7	46.1	43.7	1.5
12+ months	59	-	-	-	5	
Percent distribution	100.0	-	-	-	8.5	

* Excludes 12+ months

Table 7. Comparison of the date of accident on record and questionnaire for all households interviewed containing 1+ sample persons.

Time Lapse Between Date of Accident and Date of Interview	All Sample Households					
	Total		Reported Sample Accident		Did not report Sample Accident	
	Frequency	Percent Dist.	Frequency	Ratio Dist.	Frequency	Ratio Dist.
Total	532	100.0	453	85.2	79	14.8
Less than 3 months	106	19.9	103	97.2	3	2.8
3 - 6 months	198	37.2	177	89.4	21	10.6
6 - 9 months	101	19.0	85	84.2	16	15.8
9 - 12 months	127	23.9	88	69.3	39	30.7
Less than 6 months	304	57.1	280	92.1	24	7.9
6 - 12 months	228	42.9	173	75.9	55	24.1
Less than 12 months	532	100.0	508	85.2	79	14.8

Table 8. Comparison of the date of accident on record and questionnaire for all households interviewed by respondent status.

Time Lapse Between Date of Accident and Date of Interview	All Sample Self-responding Households						All Sample Proxy-responding Households					
	Total		Reported Sample Accident		Did not report Sample Accident		Total		Reported Sample Accident		Did not report Sample Accident	
	Frequency	Percent Dist.	Frequency	Ratio Dist.	Frequency	Ratio Dist.	Frequency	Percent Dist.	Frequency	Ratio Dist.	Frequency	Ratio Dist.
Total	287	100.0	258	89.9	29	10.1	245	100.0	195	79.6	50	20.4
Less than 3 months	54	18.8	51	94.4	3	5.6	52	21.2	52	100.0	0	0.0
3 - 6 months	107	37.3	102	95.3	5	4.7	91	37.1	75	82.4	16	17.6
6 - 9 months	55	19.2	51	92.7	4	7.3	46	18.8	34	74.0	12	26.0
9 - 12 months	71	24.7	54	76.1	17	23.9	56	22.9	34	60.7	22	39.3
Less than 6 months	161	56.1	153	95.0	8	5.0	143	58.4	127	88.8	16	11.2
6 - 12 months	126	43.9	105	83.3	21	16.7	102	41.6	68	66.6	34	33.3
Less than 12 months	287	100.0	258	89.9	29	10.1	245	100.0	195	79.6	50	20.4

Table 9. Variance and bias of the estimated probability of injury by recall period.

	RECALL PERIOD		
	Less than 3 months	Less than 6 months	Less than 12 months
Sample size	$n = 134 \times 10^3$	$n = 134 \times 10^3$	$n = 134 \times 10^3$
Probability of person being injured	$P_3 = \frac{1}{4}P_{12}$	$P_6 = \frac{1}{2}P_{12}$	P_{12}
Estimated probability of person being injured in past 12 months	$4\hat{P}_3$	$2\hat{P}_6$	$\hat{P}_{12}$
Variance of estimated probability of injury in past 12 months, based on assumption of independence	$VAR\ (4\hat{P}_3) = 16P_3 \cdot Q_3/n = \frac{4P_{12}(1-\frac{1}{4}P_{12})}{n}$	$VAR\ (2\hat{P}_6) = 4P_6 \cdot Q_6/n = \frac{2P_{12}(1-\frac{1}{2}P_{12})}{n}$	$VAR\ (\hat{P}_{12}) = P_{12} \cdot Q_{12}/n = \frac{P_{12}(1-P_{12})}{n}$
Variance due to lack of independence	$K \cdot VAR\ (4\hat{P}_3)\quad K \geq 2$	$K \cdot VAR\ (2\hat{P}_6)\quad K \geq 2$	$K \cdot VAR\ (\hat{P}_{12})\quad K \geq 2$
Proportion of persons reported as injured on motor vehicle record who reported an injury when interviewed	$K_3 = .873$	$K_6 = .788$	$K_{12} = .751$
Estimated bias of $\hat{P}_{12}$ squared = $(\hat{P}_{12} - P_{12})^2$	$(4P_3-4K_3P_3)^2 = P_{12}^2(1-K_3)^2$	$(2P_6-2K_6P_6)^2 = P_{12}^2(1-K_6)^2$	$(P_{12}-K_{12}P_{12})^2 = P_{12}^2(1-K_{12})$

Table 10. Relative root mean square error, variance and bias components for selected population sizes of injured persons.

Total Persons (in thousands)	Less than 3 months			Less than 6 months			Less than 12 months		
	Relative Root Mean Square Error %	Variance Component %	Bias Component %	Relative Root Mean Square Error %	Variance Component %	Bias Component %	Relative Root Mean Square Error %	Variance Component %	Bias Component %
25 -----	70.3	69.1	1.2	53.2	48.8	4.4	42.6	34.6	8.0
50 -----	50.5	48.9	1.6	40.5	34.5	6.0	34.9	24.4	10.5
75 -----	41.9	39.9	2.0	35.3	28.2	7.1	31.9	19.9	12.0
100 -----	36.8	34.5	2.3	32.3	24.4	7.9	30.3	17.3	13.0
125 -----	33.4	30.9	2.5	30.4	21.8	8.6	29.3	15.4	13.9
150 -----	30.9	28.2	2.7	29.1	19.9	9.2	28.6	14.1	14.5
175 -----	29.0	26.1	2.9	28.1	18.5	9.6	28.1	13.1	15.0
200 -----	27.5	24.4	3.1	27.3	17.3	10.0	27.7	12.2	15.5
300 -----	23.6	19.9	3.7	25.5	14.1	11.4	26.8	10.0	16.8
400 -----	21.4	17.3	4.1	24.5	12.2	12.3	26.3	8.6	17.7
500 -----	20.0	15.5	4.5	23.8	10.9	12.9	26.1	7.7	18.4
600 -----	19.0	14.1	4.9	23.4	10.0	13.4	25.9	7.1	18.8
700 -----	18.2	13.0	5.2	23.1	9.2	13.9	25.7	6.5	19.2
800 -----	17.6	12.2	5.4	22.9	8.6	14.3	25.6	6.1	19.5
900 -----	17.1	11.5	5.6	22.7	8.1	14.6	25.6	5.8	19.8
1,000 -----	16.7	10.9	5.8	22.5	7.7	14.8	25.5	5.5	20.0
1,500 -----	15.5	8.9	6.6	22.1	6.3	15.8	25.3	4.5	20.8
2,000 -----	14.9	7.7	7.2	21.9	5.5	16.4	25.2	3.9	21.3
2,500 -----	14.5	6.9	7.6	21.8	4.9	16.9	25.1	3.4	21.7
3,000 -----	14.2	6.3	7.9	21.6	4.4	17.2	25.1	3.1	22.0
3,500 -----	14.0	5.8	8.2	21.6	4.1	17.5	25.0	2.9	22.1
4,000 -----	13.8	5.4	8.4	21.5	3.8	17.7	25.0	2.7	22.3
4,500 -----	13.7	5.1	8.6	21.5	3.6	17.9	25.0	2.6	22.4
5,000 -----	13.6	4.9	8.7	21.5	3.5	18.0	25.0	2.4	22.6

Figure 2. Relative root mean square for aggregates by recall period.