UM-HSRI-SA-75-1

DESCRIPTION AND EVALUATION of the NEISS REPORTING SYSTEM as Related to BICYCLE-ASSOCIATED ACCIDENTS

INTERIM REPORT

to

Bicycle Manufacturers Association

1 February 1975

Jairus D. Flora, Principal Investigator
Richard J. Kaplan
Elizabeth Margoshes

Prepared By

Highway Safety Research Institute
The University of Michigan
Huron Parkway & Baxter Rd.
Ann Arbor, Michigan 48105

ABSTRACT

The National Electronic Injury Surveillance System (NEISS) is reviewed and evaluated with respect to the sampling methodology employed, the collection and management of the data, and the uses to which the results are put. Regarding bicycles specifically, previous studies in the area are reviewed, and the treatment of the NEISS reporting of bicycle-related accidents is discussed. Recommendations are made concerning the work to be done during the remaining eight months of the contract period.

TABLE OF CONTENTS

I. Evaluation of NEISS data collection system

A. Sampling

The demographic data upon which the sampling plan is based were drawn from the 1960 U.S. census. In view of the time during which the planning was done, this is quite understandable. However, there was a 13.3% increase in population in the U.S. from the 1960 to 1970 census. Further, and perhaps more important, there were geographical shifts of the population and also shifts from rural and central city to suburban dwelling. Since the place where a person lives will influence to some extent the products he uses, these differences may be important. An additional effect on the NEISS data is that the weights for the strata are no longer correct. This means that estimates formed from the data will be biased. The direction and magnitude of this bias is unknown; it may not be a major problem. However, for a system which emphasizes "real-time" data, it does seem that by now the more up-to-date census information should have been included. Some compromise between updating and maintaining the system needs to be reached. The sample cannot be changed too frequently, due to the time and effort involved in such a change. Neither can it remain the same indefinitely, however, since population shifts will eventually affect the product-associated injuries to the extent that any inferences drawn would no longer be defensible. The use of the 1960 census in the original plan does not appear to be a major drawback, but it does introduce some bias into the estimates.

The use of hospital emergency rooms as the data collection sources has some obvious drawbacks: not all persons injured are treated there, and those who are are likely to differ from those who are not. CPSC estimates

that about 40% of injuries are treated in ER, the others being treated in physicians' offices, at home, or elsewhere. To the extent that the socio-economic variables of the group treated at emergency rooms differ from the others, and to the extent that these differences influence these persons' uses of products, the data obtained from NEISS will be in error. On the other hand, ER's represent a good starting point for the collection of such data. However, these data cannot be blindly accepted as representative.

The design of the sampling plan seems to be unnecessarily complex. (See Appendix I for a detailed description of the sampling procedure.) The reasons which may have called for such a plan are not evident. I would have suggested stratifying the U.S. geographical location to ensure coverage of the entire U.S. (since product use and hence injuries are associated with geographical region and climate, and since the surveillance nature of the data desired would make it desirable to ensure coverage of products with only local distributions) and stratification by size of the institutions to improve precision. The plan currently used seems to reflect a misunderstanding of stratification: one samples from defined strata; one does not select strata at random and then sample from them. The implementation of the plan appears to have resulted in a subset of the U.S. having been selected as the population to which generalizations could be made. This subset was selected with probabilities proportional to population, so there is incomplete coverage on a geographical basis; the high-population areas were more likely to have been selected. Within this set, institutions were stratified with respect to size and location and were then sampled with probabilities related to the size of the institution. Thus, high population was in a sense given

a weight twice, making it unlikely that low-density areas would be represented. The net effect has been to eliminate large geographical areas from the sampling and to concentrate data collection mostly in densely populated areas.

The actual probabilities with which each institution (of the 2300 left in the subset of the U.S. to be sampled) was selected are extremely complicated and can only be computed with a great deal of effort and computer time. One severe drawback of this is that revision of these probabilities becomes very difficult. In the design of the sampling plan no consideration was given to the practical problem that not all of the institutions selected might agree to participate, or that some of the institutions might subsequently drop out. In actual practice, only about 60% of the hospitals selected by the sample agreed to participate. The other 40% represent second, third, or further choice replacements. Further, a number of institutions have dropped out, reducing the proportion of the original sample. The result of the nonparticipation problem is that the weights calculated for the institutions in the sample are not correct. This does not seem to have been realized by the CPSC. Further, there seems no way of calculating the correct weights. As a result, all of the projections to national estimates are in error. The magnitude of the errors is unknown. However, basing policy decisions on data which are known to be erroneous, without a solid estimate of the amount or direction of error, is certainly undesirable and should be corrected.

B. Data Management

Even if the sampling plan used by NEISS were sound, the collection and management of the data leaves much to be desired. The basic data are obtained from the individual ER records. Although this does entail the use of several different forms, this is not a serious problem so long as all of the forms contain the necessary information and so long as all are accurately recorded. However, the person who is responsible for obtaining the data is usually the admitting clerk in the ER. This is a low-skill position with a large turnover rate. Further, these persons have had little if any instruction from NEISS. Yet all the data to be obtained depends upon them. The data that they record on the ER form are the only data available to code and send to the NEISS computer.

ER personnel in general tend to be primarily interested in treating the patients, particularly if the injury is at all serious. Consequently, it is easy to imagine that if there is any rush at all, the determination of product information--which does not appear to be immediately related to the treatment--may be slighted. To the extent that this occurs, data are of poor, perhaps unusable, quality.

It is also possible for various biases to arise through this method of data generation. The CPSC has over 900 categories of products. Few persons could remember all of these with equal facility. The products which are most familiar will get the most attention in determining whether an injury was product-associated; also, any products identified by NEISS for special consideration will likely turn up more frequently than they would had they not been singled out for attention.

Assuming that there are no conscious biases on the part of those who fill out the ER records, there still remains the strong possibility of unconscious bias to record some products more frequently than others. At the very best, the basic, primary data collection is inadequately controlled, and can be expected to result in different practices at the different institutions.

Once the ER records have been completed, once a day, a person--different from the one who filled out the records initially--codes them and prepares them for transmission to the computer in Bethesda. The position of the person who does the coding varies from institution to institution, from a junior clerk to a hospital administrator. The quality of the preparation is not in question. However, in the case of severe, life-threatening injuries, the person doing the coding may recall the outcome of a case whose status has changed from the ER. For example, a severely injured person may have died, or some one who was treated and released may have returned and been admitted. If such knowledge is used in the coding, the data on injury severity will be changed.

The interactive data management via computer messages appears to be quite good. However, this merely prevents inaccuracies from entering the data during the coding or computerizing stages. It cannot help data which were wrong or missing to begin with.

The set of data elements which is collected is quite short. Consequently if does not include as much information as would be desirable. One such major gap is in the details of the injury. Injury is coded only in the NEISS matrix (see Appendix II). Further, only such injury information as

is available in the ER is usually included. Later diagnoses or developments are not included in the record unless by chance the coder happens to know of that particular case and changes the ER record. This makes it impossible to determine how a product may have caused an injury, or which injuries were associated with which products in the case of multiple product involvement.

The most crucial gap in the information, however, is the complete absence of any information on how the product was related to the injury. If such injury information is to be used to determine the relative safety of a product, then, logically, the injuries to look at are those which are in some way product-caused. Injuries caused by defects in design or construction would clearly call for remedial action on the part of manufacturer. Injuries caused by misuse, or improper use, or inadequate maintenance might have been prevented by safety information warning against the hazards of such misuse or mis-maintenance. There is a need for data which would relate an injury to the mechanism which caused it, so that preventive measures might be devised.

A final crucial element missing from the data set is information on exposure. The more persons who use a product, and the more that they use it, the more persons who will have accidents while using that product. Thus, the more popular and useful a product, the higher its raw frequency of "associated" injuries will be. An investigation would probably show that 80% or more of all injuries are "associated" with the underwear of the injured person, since probably at least this high a proportion of individuals were wearing underwear at the time of their accident. In order for meaningful comparisons to be made, a measure of risk incorporating

exposure must be used in addition to the frequencies and severities of injuries associated with products. Further, the association must not simply be a trivial one.

One other potential bias in the data collection methods needs to be pointed out. For a given injury, young children are more likely to be brought to an ER for treatment than are older children or adults. Further, a parent or guardian bringing a small child in for treatment is more likely to give adequate information on the toy or product associated with an injury than is an injured person himself. Consequently, unless some measures have been taken to correct for or remove these biases, one would tend to find young children over-represented in the ER population. Also, if this is true, one could expect the product association to be more reliable for the younger children. One would tend to see fewer product associations among older persons. One would also see a more severe distribution of injuries among older persons.

II. Use of NEISS Data

The NEISS data, as described in the foregoing sections of this report, are used by CPSC as an indicator of the safety performance of the products under their jurisdiction. The Commission maintains that the function of these data is limited to this indicator role, in that the data only provide a direction for further, more detailed investigations on their part. The publication of the results of this data collection effort, however, tends to present them more with the status of "facts" rather than indices for internal consumption, and might be construed as somewhat of an inflation of the meaningfulness of the system output. The remainder of this section discusses the process by which the reported statistics are derived.

A. Frequency-Severity Estimates

CPSC deemed it necessary, for purposes of cross-product comparisons, to have a single number represent the degree to which a specific product was responsible for the cause of injury in the population. The result of this assumption is the production of a frequency-severity index--that is, the sum for each product of the number of injuries, each multiplied by its severity. Expressed in more mathematical terms:

$$I_{FS} = \sum_{i}^{n} S_i$$

where I_{FS} is the frequency-severity index for a specific product. S_i is the severity attached to the i^{th} injury, and n is the number of reported injuries for the product.

While having a considerable amount of face validity, such an index has

obvious limitations. First, the frequency involved is used to generate an estimate of the nationwide index for the occurrence of such injuries and is, therefore, subject to the effectiveness of the sampling plan in achieving representativeness. This has been discussed above. Secondly, as will be discussed below, the severity index itself is an arbitrary one, and there is no apparent justification in combining these in the form of a linear additive model. The presumption, for example, that an injury at leval 4 is three times as bad as an injury at level 1 cannot be accepted without some rationale in terms of such variables as cost, discomfort, disability over defined time periods, days of lost work or lessened efficiency, and the like. Toward the higher end of the injury spectrum, the linear additive assumption in combining the data elements appears even more tenuous.

B. Age "Adjustment"

The CPSC, feeling, and probably correctly so, that the same extent of injury affects various age groups differentially, thought it wise that the indices produced by NEISS be adjusted or weighted for age. The original decision was to give a weighting factor of two to injuries occurring to individuals ten years old or younger and to those sixty-five years of age and older. The weighting has subsequently been removed from the data pertaining to the elder of these two groups.

This weighting scheme is arbitrary; no rationale is given for using it other than some obvious common-sense criteria. As to the younger group, it is assumed that CPSC is concerned, as is our society in general, with

adequately protecting children. The weight then is present to help identification of products which are more dangerous to children. It is well known, however, that the younger group will suffer less injury and recover faster from the same trauma than will an older group, and it is nowhere evident that a weight based on age alone, rather than a combination of age and type of injury, will do anything toward making a useful adjustment. In the case of the older victims, the reverse is true. The same trauma may well produce injuries which are both more severe and more costly to the individual and to society, and yet the attempt to get at this phenomenon by an increased weight was discarded.

To the credit of CPSC, we must say that they are well aware of the inadequacies of the scheme currently in use with regard to weight. Their staff is presently at work on new and hopefully better procedures, and the above critique is intended to emphasize the complexity of the problems they face.

C. Scheme for Weighting Severity

The operators of NEISS are also the prime users of the system, and one of their chief objectives is to detect problems with specific products. The frequency of associations of products with injuries is not considered alone to be a good enough indicator of a problem area. The severity of injury would affect their consideration, and since more severe injuries would indicate more severe problems, it was decided to weight the severe injuries more heavily than less severe ones. The weighting scheme used is purely arbitrary, and the NEISS people are quoted as saying that it

"has an algebraic rationale but not a real-life rational." It seems unlikely that an arbitrary weighting scheme can be said to have an "algebraic rationale," but what this is intended to imply is that the formula for calculating the weights is constant. After arbitrarily assigning a weight of ten to a severity designation of 1, the other weights are calculated by:

$$W_i = W_{i-1} + (.10)(2^{i-1})(W_{i-1})$$

where W_i is the weight of the i^{th} severity index. This formula produces the following table:

1	10
2	12
3	17
4	31
5	81
6	340
7	2516
8	34721

The last entry in the table has recently been revised, and the weight for the highest severity index has been assigned, again arbitrarily, the value of 2516.

Without a great deal of further research on the topic, nothing substantial can be said in defense of or in opposition to the severity weighting currently in use. It is simply one of an infinite number of such schemes. It is not, however, at all obvious to the general reader of the NEISS statistics how the weighting may have affected the numbers presented. We intend, in a later portion of this program, to develop alternative weighting schemes, each based on a set of definable criteria, and recalculate what the NEISS

output would be, given each of these. We would hope thereby to gain some insight into the complexity of the problem and into the sensitivity of the output to weighting differences.

D. In-Depth Studies

Although this report deals very heavily with the NEISS data collection and reporting system, it must be recognized that this system represents only the first step in the process of performing CPSC's function. No decisions as to regulatory actions the commission is empowered to take would be made on the basis of NEISS data alone. NEISS has the function of sensitizing the commission staff to problems which may exist. Further steps taken before the institution of any regulatory activity include: (1) product testing in either their own or independently operated laboratories; (2) consultation and discussion with the product manufacturers; and (3) more detailed studies of specific cases identified by NEISS as potentially important.

The last of these, the so-called "In-Depth" studies, since they are a subset of the cases occurring in the NEISS data set, are restricted in their efficiency by the limitations of NEISS itself. They are considered necessary, however, because it is only by such studies that the actual causes of accidents can be determined. Among the causes of product-associated accidents are defects in design or manufacture, inadequate or misleading information on product use, improper or non-intended use of the product, or merely the proximity of the product at the time of the accident. Each of these possibilities, or a combination of them, implies a different type

of action to be considered on the part of the commission, and this information as to causality is considered vital to their deliberations.

Any evidence acquired through in-depth studies of NEISS cases, while useful for producing insights into possible problems, must at this time be viewed as rather weak from the statistical point of view. The reason for this is, first of all, the sampling deficiencies of the NEISS system itself. Beyond that, however, there appears to be no systematic procedure defined for the selection of cases to be studied in depth, and the studies are restricted to those where the voluntary cooperation of the accident victims is readily available. Time and budget contraints make it difficult for CPSC to do otherwise at this time, but if this class of activity is to attain a status other than as a stimulator to intuition, considerable thought must be given to its proper design and execution.

III. Applications of NEISS Data to Bicycle Safety

Several deficiencies in the sample design and the data collection and management of NEISS by the CPSC have been discussed in the preceding sections. A number of these deficiencies seem to apply with particular force to uses of the NEISS data in considering bicycle accidents.

As mentioned before, the error in the sampling scheme resulted in the selection of a complicated geographical subset of the U.S. as the population that was in fact sampled. This subset is roughly the drawing areas for the 2300 hospitals which were not eliminated at the first stage. Since the elimination was based on probabilities which were proportional to population, the chosen subset is presumably of higher population density than that of the U.S. as a whole. In viewing a map with the NEISS hospitals located on it, it can be seen that large regions of the Midwest and West (other than the West Coast) have been excluded. Thus the regions which were in fact sampled would have been those with high traffic density, and could be expected to have rather more bicycle-traffic accidents than the more rural, small-town areas.

A great deal of the bicycle use in this country is by children. Accidents which result in minor injuries to children are more likely to result in medical attention than are the same injuries to adults. As a consequence the ER rates will typically be more complete for younger children than for adults. Hence products which are used more by children than by adults may be expected to appear in somewhat excessive proportions. When coupled with "age-adjustment" to severity suggested by CPSC, this appears to present a strong bias toward over-representation of young groups.

The calculation of the so-called "consumer product hazard index" and its publication by the CPSC does not appear to be an appropriate use of the surveillance data from NEISS. First, the NEISS data do not appear to be sufficiently reliable for such use to be made of them, due to the sampling and data collection inadequacies. Second, the index is based on an entirely arbitrary injury severity scale. Third, the index is based on mere association--there are no data to indicate the presence or degree of product fault in any of the injuries. Fourth, the index is simply a sum of raw data; the more popular and widely used a product is, the larger its score on this index will be. In order to construct a "hazard index" one must have and incorporate some sort of exposure data to indicate any sort of relative risks of products.

An index of the sort used by the CPSC can be useful as an internal device to alert the CPSC to areas where there are substantial numbers of accidental injuries. However, to some extent it does this by pinpointing products which are widely used. It does not pinpoint products which are particularly hazardous to the user. In order to assess the hazard of a product, more information is needed. First, there is a need for exposure data. This need seems to have been somewhat slighted by CPSC. Secondly, there is a need for additional information on the types of injuries and their relation to the product in question. It is crucial to determine whether the accident and/or resulting injuries were the result of some cause unrelated to the product in question, or were in fact the result of the product. If they were actually related to the product, then the possibility

of some action to reduce the risk or severity of such injuries is present. However, there is a need for sufficient information to determine whether the cause was due to a defect in design or construction, which would call for a construction standard; or to inadequate maintenance, which might call for an educational standard; or due to intentional misuse of the product, which might call for a warning standard, or perhaps for legislation dealing with the use of the product (enforce the prohibition against riding a bicycle the wrong way on a one-way street, for example).

In order to obtain data of sufficient detail to use in these aspects of the CPSC, in-depth investigations are necessary. The cases for such investigations must be selected and pursued in such a manner that the results can be related to the frequency estimates for the whole set of accidents. Thus, in order to recommend a bicycle standard, information should be available to indicate how it would help prevent injuries in certain types of accidents, and also what the frequency of such accidents is. It is impractical to write a standard which could apply, on the average, to only one accident a year.

IV. Literature Results

Several studies in the last five to six years have attempted to draw conclusions about the nature of bicycle accidents occurring, the characteristics of the population to which they occur, and/or the properties of bicycles involved in accidents. They have ranged in scope from a single city's children to a state's bicycle-motor vehicle accidents over three years involving children in five nationally distributed semi-urban areas. F.J. Vilardo and J.H. Anderson (1969) gathered exposure and accident data on 20,000-plus children in grades 2 through 8 in Arlington Heights, Illinois; Ann Arbor, Michigan; East Baton Rouge Parish, Louisiana; Los Angeles, California; and the State of Delaware. The sex and age distributions of all bicyclists and the approximately 4,000 experiencers of bicycle accidents were fairly similar. The per cent distribution of ostensible accident causes is shown in Table (a)-2. Vilardo and Anderson observed that boys had more motor-vehicle-involved accidents and girls more falls (riders of girl's bicycles experienced "significantly more knee injuries"). In approximately 70 per cent of the accidents the bicycle had been used as a toy, in 23 per cent as a general transportation vehicle, and in 7 per cent as a transportation vehicle specifically to or from school. While 38 per cent of exposure respondents claimed night-time riding, 3 per cent of accidents occurred then.

E. Brezina and M. Kramer (October 1970) looked at 275 bicyclists (5-14 years, 93% males) in reportable bicycle/motor-vehicle collisions in the 10-month school 1969-70 and a randomly chosen comparison group of 1,000-plus 8- to 13-year-old male bicycle owners. According to the investigators, risk factors

"related primarily to the bicyclist's comprehension of risk inherent in different roadway environments, prediction of traffic movement, and control of the bicycle.... Among the bicycle factors examined, the adjustment of the bicycle to the rider has the most significant influence on the risk of collision, especially among young riders." Relative risk of collision involvement increased as bicyclist ground-level clearance increased. Approximately half the collisions involved standard configuration bicycles, while 35 per cent of the comparison group owned such bicycles. The investigators, lacking use information by configuration type, were reluctant to draw firm conclusions regarding relative risk but they suggested "among 11- to 13-year-olds a lower collision involvement rate... is found for high-rise bicycles" compared with standard configuration bicycles. Further, they suggested younger riders of hand-brake-equipped bicycles are more at risk than those of bicycles with foot brakes [note that Rice and Roland (April 1970), studying the dynamics of bicycle design, suggest front wheel (hand) brakes have "hazard potential"]. 55 per cent of 259 collisions involved bicycles crossing some flow of traffic; 57 per cent occurred within one block of home.

In February, 1969, P. A. Waller and D. W. Reinfurt published an investigation of some 2,400 fatal and non-fatal bicycle/motor-vehicle accidents reported in North Carolina July 1965 to June 1968 relative to 1966 motor vehicle accidents. They analyzed the physical circumstances surrounding the accident, driver and bicyclist characteristics, and vehicle (car) and road characteristics. The authors summarize: "It appears that the typical bicycle accident occurs in clear dry weather during daylight hours....The

cyclist is usually a young male between 10 and 14 years of age who apparently emerges unexpectedly from....[an] intersection of some sort. Fatal bicycle accidents appear to be associated with the older bicycle rider." About 25 per cent of bicycle fatalities occurred at night on unlighted roads, but only 6 per cent of all bicycle accidents occurred under such conditions.

A four-month study of ownership, use, and injury patterns among 3- to 12-year-olds in the Burlington, Vermont area (J. A. Waller, 1970) was intended to determine if standard and high-rise bicycles differ in associated rate and/or severity of injuries. With 104 injuries (97% of all occurring) and 6,200 comparison cases, no differences were observed between the two bicycle styles. A per cent distribution of accident causes is shown in Table (a)-3. While about 30 per cent of owners stated they rode after dark, no accidents occurred then. Eight accidents involved contact with an automobile; 69 per cent occurred within one block of home. In the range 5 to 12 years, of 6,185 owners a four-month injury rate of 11.2 per 1,000 owners (based on a total of 69 injuries) was calculated.

During approximately the same period as the Vermont study, exposure and accident data (the former from a school sample, the latter from the sample plus emergency room and police department records) for children 5 to 18 were gathered in Raleigh, North Carolina. E. A. Pascarella (1971) found no real differences in frequency or category of accident associated with bicycle style [a breakdown of the latter may be found in Table (a)-1]. Accident subjects tended to be younger and to have fewer years riding experience than those without accidents. Bicycle age and condition and passenger-carrying status did not appear to influence risk... An overall accident rate of 1.58 per 1,000 miles (based on 60,109 miles, 95 accidents

and 397 owners) was calculated. It was further estimated that 15.7 per cent of accidents would result in no injury, 76.3 per cent in mild injury (first aid only), and 8 per cent in physician-, emergency room-, or other hospital-treated injury. [Contrast this distribution to the Market Facts, Inc. survey (1970) cited by NEISS that says 82 per cent of injuries are treated professionally.] Pascarella uses ownership, mileage, and accident rates to project approximately 9.46 million accidents occurring annually in the United States, with about 760,000 requiring medical attention. Note that this estimate, made in 1971, is nearly twice that made by NEISS (403,000) in 1973.

NEISS conclusions about the association of accidents with any particular aspect of bicycle style, structure, or condition is made primarily from their "in-depth" investigations rather than reports on all accident cases arriving at emergency rooms. These in-depth cases have included referred death certificate information as well as any very severe (category 6, 7 or 8) injuries and cases that look interesting or are called to someone's attention. The surveillance data give frequencies of injury types and to which age and sex groups. For example, of 11,403 injuries during the period July 1972 - March 1973, 13 per cent were fractures (J. A. Waller reported 20 per cent fractures out of 104 injuries in Vermont). Using NEISS in-depth cases, a per cent distribution of accident causes is shown in Table (a)-4. The age and sex distributions of non-fatal cases and of total input records from the May, 1973, Staff Analysis are virtually the same.

Table (a)

	1. Pascarella			2. Vilardo		
	All types	High-rise	Std	All types	Middle-weight	High-rise
N	218			3,952		
All Types	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%
Bicycle Struck Car	2.8	1.9	3.5	7.9	8.4	6.0
Car Struck Bicycle	6.4	3.7	8.2	5.3	5.3	4.4
Bicycle Struck Fixed Object	4.1	3.7	5.9	15.1	14.5	16.0
Bicycle Struck Bicycle	7.3	8.3	4.7	18.3	17.4	19.3
Fall	72.1	75.0	69.4	53.3	54.3	54.3
Other	7.3	7.4	8.3	-	-	-

	3. J. A. Waller			4. NEISS	
	All types	High-rise	Std	All types	
N	104			119	
All types	100.0%	100.0%	100.0%	100.0%	
Car Involved	9.6	2.0	14.0	-	
Hit Obstruction	17.0	15.0	19.0	-	
Loss of Control	20.0	22.0	18.0	63.2	
Foot Caught	11.0	10.0	11.0	10.8	
Bike Broke	2.0	2.0	3.0	16.7	Body Entanglement Mechanical and Structural
Rode too close to Stationery Object	8.7	10.0	8.0		
Other	31.7	39.0	27.0	9.2	

Table (b)

Breakdown of "Mechanical and Structural"		
	Pascarella	NEISS
N	30	20
All types	100.0%	100.0%
Brake Failure	20.0	35.0
Wheel and/or Handlebars Unstable or Disengaged	50.0	35.0
Other	30.0	30.0

	(estimates) Accident Facts			P.Waller & Reinfurt	J.Waller	Idaho	Vilardo	Pascarella
	1973	1971	1969					
Total Comparison Bicyclists	N.A.	N.A.	N.A.	N.A.	6,185	N.A.	20,190	1,204
Total Bicycle Accidents	10^6	N.A.	N.A.	N.A.	107	N.A.	3,952*	217
Total Bicycle/Motor-Vehicle Accidents	41,150	40,850	39,820	2,453	8	467	520	20
Total Fatalities	1,250	N.A.	N.A.	N.A.	N.A.	N.A.	2	N.A.
Total Bicycle/Motor-Vehicle Fatalities	1,150	850	820	105	N.A.	12	N.A.	N.A.
Period of Study	1 yr.	1 yr.	1 yr.	3 yrs.	4 mos.	3 yrs.	2 wks.	5 mos.
Age Range	All	All	All	All	3-12	All	7-14	5-18
Date of Study	1973	1971	1969	1970	1970	1973	1969	1970
Area	U.S.	U.S.	U.S.	N.C.	Vt.	Id.	5 cities	Raleigh, N.C.

*Respondents were asked if they ever had an accident.

V. Discussion and Recommendations for Further Analyses

The authors met with the managers of the NEISS system at CPSC in October of 1974 and at that time requested that their data tapes be made available to HSRI. The tapes for 1972 through 1974 are now in hand and are in the process of being readied for use with MTS (the Michigan Terminal System operating on an IBM 370/168 computer).

We propose to use the NEISS data in four basic research areas as follows:

(1) Extract data on individual products in a manner which decomposes the frequency-severity index into its component parts. This will allow us an unweighted view of the accident distribution over products for each level of severity coded and provide the basis for the next step.

(2) Determine the sensitivity of the NEISS output to differences in both age and severity weighting schemes. We will construct a number of reasonable weighting schemes for both age and injury severity and generate for a selected list of products the resulting frequency-severity indices. This will allow us to determine the effect of these weights as an influence in the possible reordering of priorities which are designated by the results.

(3) Establish the influence of errors in sampling on the projection of nationwide estimates. By a reassignment of the probability fractions associated with the different sampling units we can investigate the effect of taking into account, for example, changes in the census data from 1960 to 1970.

(4) Calculate the estimated variance in the frequencies of accidents estimated for a selected list of products. With this variance estimate and

the bias estimate discussed in (3), we can estimate the mean square error in the frequency projections. Such information is necessary, since it reflects the adequacy of the data and the magnitude of errors which may result if policies are based on the data.

HSRI has and maintains a large set of data files covering highway accidents over a large geographical area and a substantial time period. It is our intention to search these files with respect to automobile/bicycle accidents to establish a basis for comparison with bicycle accidents alone. We will seek also to determine if other data sets exist which will help us in any way in establishing the nature and extent of the bicycle accident problem.

REFERENCES

Blake, R., Jr. "Pedestrian and Bicycle Accidents." September 1973, State of Idaho Department of Highways Planning of Traffic Division.

Brezina, E. and Kramer, M. "An Investigation of Rider, Bicycle and Environmental Variables in Urban Bicycle Collisions." Technical Bulletin SE-70-01, October 1970, Ontario Department of Transport.

National Safety Council. "Accidents Facts." 1970, 1972, 1974.

National Transportation Board. "Special Study: Bicycle Use as a Highway Safety Problem." Report No. NTSB-HSS-72-1, April 5, 1972.

Pascarella, E. A., et al. "Characteristics of Youthful Bicycle Riders in an Urban Community and Events Accruing to Operation of their Vehicles." June, 1971, University of North Carolina Highway Safety Research Center.

Rice, R. S. and Roland, R. D., Jr. "An Evaluation of the Performance and Handling Qualities of Bicycles." Technical Report No. VJ-2888-K, April 1970, Cornell Aeronautical Laboratory, Inc.

United States Bureau of the Census. 1970 Census of the Population: Number of Inhabitants, United States Summary. Vol. 1, no. 1 A. December 1971.

Vilardo, F. J. and Anderson, J. H. "Bicycle Accidents to School Aged Children." Report No. 169, September 1969, National Safety Council.

Vilardo, F. J. et al. "An Investigation into Bicycle Usage." Report No. 268, September 1968, National Safety Council.

Waller, J. A. "Bicycle Ownership, Use and Injury Patterns Among Elementary School Children in Chittenden County, Vermont." August 1970.

Waller, P. A. and Reinfurt, D. W. "Bicycles: An Analysis of Accidents in North Carolina." February 1969, University of North Carolina Highway Safety Research Center.

Linda Vanty Soley

Appendix I

NEISS SYSTEM A ESTIMATOR EQUATIONS

A report to the

Division of Consumer Product Safety

Bureau of Product Safety

Food and Drug Administration

PHS, DHEW

by

Thomas D. Downs

Associate Professor of Biometry

The University of Texas At Houston

School of Public Health

P.O. Box 20186

Houston, Texas 77025

November , 1971

TABLE OF CONTENTS

NEISS SYSTEM A ESTIMATOR EQUATIONS

1. INTRODUCTION

Purpose. The purpose of this document is to portray the mathematical equations for national estimates of consumer product injuries, based on NEISS System A data.

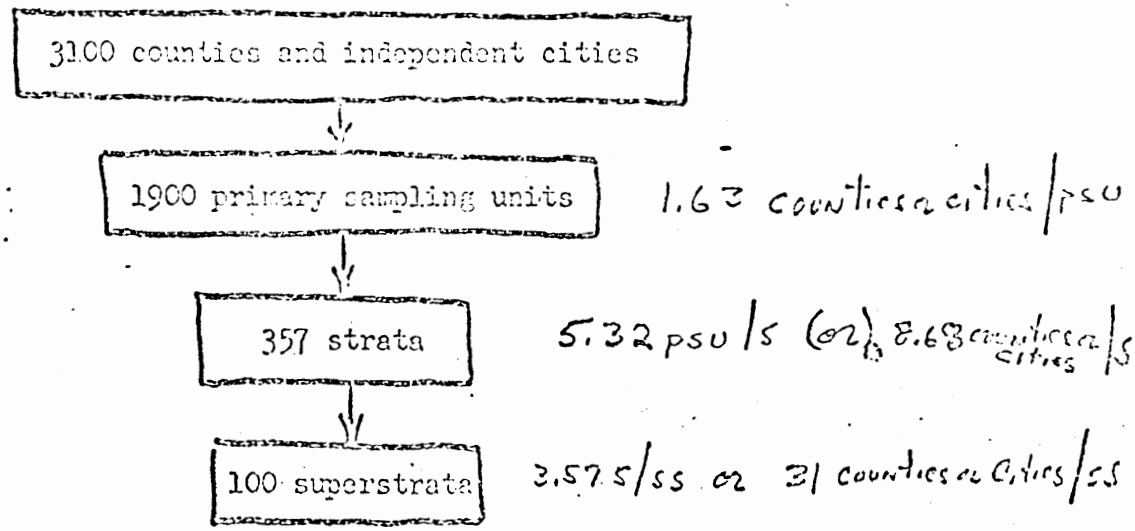
The primary responsibility of the Division of Consumer Product Safety is to protect the American consumer against unduly hazardous or potentially hazardous products used in and around the home or in recreational areas. To fulfill this responsibility the Division of Consumer Product Safety must first determine how many of such injuries occur, the severities of these injuries, and the types of products involved. The estimator equations permit national estimates of various items of interest, such as the number of injuries of a specific nature. These national estimates, in turn, may be used to provide program direction, such as:

1. Identify the magnitudes of these injury problems.
2. Identify specific consumer products directly related to such injuries.
3. Suggest the need for standards, safety actions, or remedial actions.
4. Provide necessary data for evaluating the effectiveness of those items in 3, where action has been taken.
5. Assist in planning future improvements to the NEISS System A.

Background. The Division of Consumer Product Safety is currently implementing a National Electronic Injury Surveillance System (NEISS). The data collection of injuries by means of querying a probability sample of

hospital emergency units is designated as System A within NEISS. The sampling plan used is similar to that for the Health Interview Survey (HIS), which itself is similar to that of the Current Population Survey (CPS)[1].

For sampling purposes the CPS constructed the following hierarchical structure for the United States: There are some 3100 counties and independent cities in the U.S.. These were combined in various ways to form 1900 contiguous units called primary sampling units. These 1900 were subsequently combined to form 357 strata. Finally, the 357 strata were combined to form 100 superstrata. The sampling plan of the CPS (and also the HIS) is built around this hierarchical structure. A schematic representation of the structure is given below:



[1] U.S. Bureau of the Census: The Current Population Survey--A Report on Methodology. Technical Paper No. 7. Washington, D.C. U.S. Government Printing Office, 1963.

2. THE SAMPLING PLAN

The NEISS System A sampling frame consists of those hospitals in the U.S. which had emergency units in 1963, as determined by the American Hospital Association at that time. Long-term hospitals and federal penal hospitals are excluded from the frame. After these exclusions, there remain 4906 hospitals with emergency units, and it is from these 4906 that the sample of hospitals is to be drawn.

The sampling occurs in 2 stages. Stage I consists of selecting one stratum from each of the 100 superstrata. This yields 100 strata, and Stage II consists of sampling from within these 100 strata the actual individual hospitals to be included in the survey.

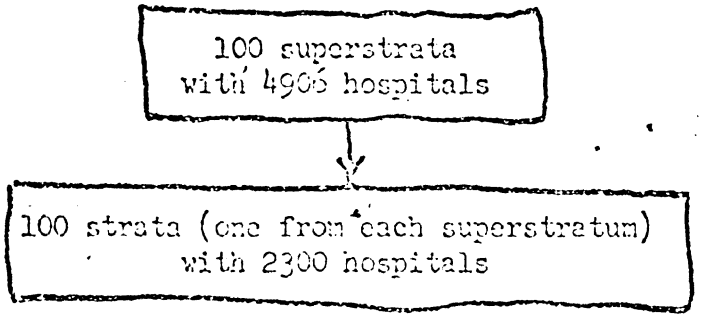
NEISS: Stage I. For each of the 100 superstrata, the stratum chosen to represent it by Stage I is selected at random on a population proportional basis. The probability of selecting a particular stratum within a superstratum is equal to

$$\frac{\text{human population* of the particular stratum.}}{\text{human population* of the superstratum}}$$

For example, if a superstratum consists of 3 strata whose populations are 1000, 2000, and 3000, then the probabilities that the first, second, or third of these would be selected to represent the superstratum would be 1/6, 2/6, and 3/6, respectively.

* Human population is based on the 1960 U.S. Census

Of the 4906 hospitals in the sampling frame, 2606 appear in the 257 strata eliminated by Stage I, and the remaining 2300 appear in the 100 strata determined by Stage I. Subsequent sampling (Stage II) is confined to sampling from these 2300 hospitals. A schematic presentation of Stage I is given below:



NEISS: Stage II. Stage II consists of sampling, from the 2300 hospitals determined by Stage I, the actual individual hospitals to be included in the survey.

A useful method of increasing the precision of injury estimates is to stratify the hospitals in the sample frame into homogeneous units called strata. Then at least two hospitals are sampled from each stratum. For NEISS System A stratification was done on a geographical and hospital-size basis. This process yielded 39 strata internally homogeneous with regard to geographical location and hospital size.

The geographical portion of the stratification was done by combining the 100 strata determined by Stage I into 13 Blocks of approximately equal human population. The Blocks are relatively homogeneous geographically. Each Block is completely contained in one of the four regions of the U.S. (Northeast, North Central, South, of West).

The stratification by hospital-size was done by using the number of emergency room visits (ERV's) in 1968 across hospitals. To this end, the

number of ERV's for each of the 2300 hospitals during 1968 was determined from existing records. When these were not reported the number of ERV's was estimated according to defined rules. Then the hospitals within each of the 13 Blocks were ranked by their determined number of 1968 ERV's. Next each Block was decomposed into Subblocks 1, 2 and 3 as follows:

For each Block, Subblock 1 for that Block was defined as the initial group of hospitals on the rank ordered listing accounting for 50% of the total number of ERV's within the Block. The group of hospitals accounting for the next 30% was denoted as Subblock 2. The group of smallest hospitals, accounting for the remaining 20%, was called Subblock 3. This process yielded the 39=13 x 3 strata.

As an example of Subblock determinations, suppose a Block had 10 hospitals within it, as determined by Stage I, and the rank ordering of these hospitals by number of ERV's was as follows:

Hospital	H_1	H_2	H_3	H_4	H_5	H_6	H_7	H_8	H_9	H_{10}	Total
No. of ERV's	10	8	7	6	5	4	4	3	2	1	50
Subblock		1			2			3			

Subblock 1 consists of the largest hospitals accounting for 25 ERV's (50% of 50), Subblock 2 accounts for 15 ERV's (30% of 50), and Subblock 3 accounts for 10 (20% of 50).

The actual distribution of the 2300 hospitals determined in Stage I, by Block and Subblock, is given below (Table 1).

TABLE 1

Distribution of the 2300 Hospitals determined in Stage I,

by Block and Subblock

Block		Subblock 1	2	3	Total
1	11	28	55	145	228
2	12	38	51	110	199
3	13	21	29	95	145
4	21	32	56	130	218
5	22	18	30	89	137
6	23	19	27	119	165
7	24	21	30	112	163
8	31	16	22	63	101
9	32	5	16	108	129
10	33	16	33	111	160
11	34	16	46	145	207
12	41	27	56	174	257
13	42	29	47	115	191
Total		280	493	1518	2300

The 39 cells in Table 1 are the final strata used for MEISS System A. All estimates are based on these 39 strata.

It remains to determine the number of hospitals to be sampled within each stratum, and to specify the methods for selecting them. The sampling rates are, for each block, about 18% for Subblock 1, 7% for Subblock 2, and 2% for Subblock 3. Subblocks with larger hospitals are sampled at greater rates because the larger hospitals have more variation. The percent of total ERV's sampled will be about 11½% for the 2300 hospitals:

$$11\tfrac{1}{2}\% = \tfrac{5}{10}\,18\% + \tfrac{3}{10}\,7\% + \tfrac{2}{10}\,2\%.$$

The total number of hospitals sampled is 119, and their distribution, by Block and Subblock, is given in Table 2.

The 5 hospitals shown in Table 2 for Subblock 1, Block 1, for instance, are randomly selected without replacement from the 28 hospitals for the same stratum in Table 1. Roughly speaking, the probabilities of selection are population proportional on a 1968 ERV basis.

TABLE 2

Distribution of the 119 Hospitals sampled in Stage II,

by Block and Subblock

Block	Subblock 1	2	3	Total
1 11	5	4	3	12
2 12	6	3	2	11
3 13	4	2	2	8
4 21	5	4	3	12
5 22	3	2	2	7
6 23	3	2	3	8
7 24	4	2	2	8
8 31	3	2	2	7
9 32	2	2	2	6
10 33	3	2	2	7
11 34	3	3	3	9
12 41	5	4	4	13
13 42	5	3	3	11
Total	51	35	33	119
	18%	7%	2%	

3. PROBABILITY CALCULATIONS

Data Required. Consider an arbitrary cell from the 39 cells in Table 2, and an arbitrary hospital, H_x, in that cell. A general expression will be derived for the probability, $P(H_x)$, that this hospital is included in the sample of 119 hospitals. This general expression may then be used, at least in theory, to calculate such probabilities for each of the 119 hospitals in the sample. To calculate $P(H_x)$ we require:

1) s=the 1960 human population of the CPS Stratum containing H_x.

2) S=the 1960 human population of the CPS superstratum containing H_x.

3) m=the number of hospitals (Table 1) in the cell containing H_x.

4) r=the number of hospitals sampled (Table 2) in the cell containing H_x.

5) A rank ordering of the m hospitals by number of ERV's:

Hospital: $H_1, H_2, \ldots, H_m$.
No. of ERV's: $n_1 \geq n_2 \geq \ldots \geq n_m$.

Items 3), 4) and 5) are common to every hospital in the same cell of Table 2, but items 1) and 2) are not necessarily so.

Probability Formulas. With the above notation, r hospitals are randomly selected, without replacement, from the m hospitals in the cell, these m hospitals being determined by Stage I. The probability that H_x is selected is the product of a) the probability that the CPS stratum containing H_x is selected in Stage I, and b) the probability that H_x is selected as one of the r hospitals in Stage II, given that the CPS stratum containing H_x was selected in Stage I.

The probability in a) is simply c/S. For b), H_x can be selected first, second, ..., or last. Let p_1 be the probability that H_x is selected first, p_2 the probability H_x is selected second, etc. to p_r being the probability H_x is selected last. The probability for b) is then $p_1+p_2+\ldots+p_r$, and thus

$$P(H_x) = \frac{c}{S}(p_1+p_2+\ldots+p_r). \qquad (0)$$

Let $n=n_1+n_2\ldots+n_m$ be the total number of ERV's for the m hospitals in the cell. Then the probability that H_x is chosen first is $p_1=n_x/n$. The probability that H_x is chosen second is

$$p_2 = \sum^{*} P(\text{a different hospital first})P(H_x \text{ second, given the first hospital}),$$

where the * indicates the summation is to be done over all hospitals except H_x, since H_x is to be chosen second and sampling is done without replacement. Thus

$$p_2 = \sum^{*} \frac{n_i}{n} \frac{n_x}{n-n_i},$$

the summation being over the (m-1) i's different from x. This last equation may be rewritten as

$$p_2 = \frac{n_x}{n} \sum^{*} \frac{n_i}{n-n_i}.$$

Similarly, the probability that H_x is chosen third is

$$p_3 = \frac{n_x}{n} \sum^{*} \frac{n_i}{n-n_i} \frac{n_j}{n-n_i-n_j},$$

the summation being over all i and j (from 1 to m) for which $i\neq x$, $j\neq x$, and $i\neq j$. Thus there are (m-1)(m-2) terms to be summed, no two of the indices i,j,x being equal in any term.

According to Table 2 r is at most 6, so only p_1, p_2, ..., p_6 require mathematical expression. These formulas are:

$$p_1 = \frac{n_x}{n} \tag{1}$$

$$p_2 = \frac{n_x}{n} \sum_{i}^{*} \frac{n_i}{n-n_i} \tag{2}$$

$$p_3 = \frac{n_x}{n} \sum_{ij}^{**} \frac{n_i\, n_j}{(n-n_i)(n-n_i-n_j)} \tag{3}$$

$$p_4 = \frac{n_x}{n} \sum_{ijk}^{***} \frac{n_i n_j n_k}{(n-n_i)(n-n_i-n_j)(n-n_i-n_j-n_k)} \tag{4}$$

$$p_5 = \frac{n_x}{n} \sum_{ijk\ell}^{****} \frac{n_i n_j n_k n_\ell}{(n-n_i)(n-n_i-n_j)(n-n_i-n_j-n_k)(n-n_i-n_j-n_k-n_\ell)} \tag{5}$$

$$p_6 = \frac{n_x}{n} \sum_{ijk\ell m}^{*****} \frac{n_i n_j n_k n_\ell n_m}{(n-n_i)(n-n_i-n_j)(n-n_i-n_j-n_k)(n-n_i-n_j-n_k-n_\ell)(n-n_i-n_j-n_k-n_\ell-n_m)} \tag{6}$$

each summation being over all possible permutatations of the indices. Thus p_1 has 1 term, p_2 has m-1, p_3 has (m-1)(m-2), p_4 has (m-1)(m-2)(m-3), p_5 has (m-1)(m-2)(m-3)(m-4), and p_6 has (m-1)(m-2)(m-3)(m-4)(m-5) terms. Equations (2) through (6) above hold for every cell of Table 2, regardless of what r is. For any particular cell of Table 2 though, where r hospitals are sampled, only p_1 through p_r need be calculated, as per equation (0).

Example. Suppose for simplicity that there are m=4 hospitals in a certain cell, and r=3 of them are to be sampled. Suppose the data for the 4 hospitals are as follows:

	Hospital			
data	H_1	H_2	H_3	H_4
s	3	4	5	6
S	10	20	30	40
n_i	4	3	2	1 , n=10

We shall find the probability, $P(H_x)$, that H_x is in the sample when x=4:

$$s/S = 6/40$$

$$\rho_1 = \frac{n_4}{n} = \frac{1}{10}$$

$$\rho_2 = \frac{n_4}{n}\left(\frac{n_1}{n-n_1} + \frac{n_2}{n-n_2} + \frac{n_3}{n-n_3}\right) = \frac{1}{10}\left(\frac{4}{6} + \frac{3}{7} + \frac{2}{8}\right) = \frac{113}{840} \sim 0.1345$$

$$\rho_3 = \frac{n_4}{n}\left[\frac{n_1}{n-n_1}\left(\frac{n_2}{n-n_1-n_2} + \frac{n_3}{n-n_1-n_3}\right) + \frac{n_2}{n-n_2}\left(\frac{n_1}{n-n_2-n_1} + \frac{n_3}{n-n_2-n_3}\right) + \frac{n_3}{n-n_3}\left(\frac{n_1}{n-n_3-n_1} + \frac{n_2}{n-n_3-n_2}\right)\right]$$

$$= \frac{1}{10}\left[\frac{4}{6}\left(\frac{3}{3} + \frac{2}{4}\right) + \frac{3}{7}\left(\frac{4}{3} + \frac{2}{5}\right) + \frac{2}{8}\left(\frac{4}{4} + \frac{3}{5}\right)\right] \sim 0.2143$$

$$P(H_x) = \frac{s}{S}(\rho_1 + \rho_2 + \rho_3) = \frac{6}{40}(0.1000 + 0.1345 + 0.2143) \sim 0.0673$$

Computer Calculations. The number of terms summed in the calculation of p_r, for a hospital H_x in a cell of Table 1 with m hospitals, is (m-1)(m-2) ...(m-r+1). This is the number of permutations of m-1 things taken r-1 at a time, and is approximately m to the power r-1. This number can be reduced by a factor of (r-1)(r-2)...3.2.1 by expressing formulas (1) through (6) as summations over combinations rather than permutations. The appropriate formulas, (1') through (6'), are given below, and their derivations appear in the appendix. A computer program was written to calculate p_1 through p_r and this, together with a documentation for it, also appears in the appendix. The computer program was checked on a hypothetical cell of Table 2 where m=20, r=6, and the numbers of ERV's for the 20 hospitals were in the ratios

1:2:3:4:...:18:19:20

A second check was done with m=20 and r=6 also, but with the hospital ERV's now in the ratios

21:22:23:24:...:38:39:40

The program was written in Fortran IV and ran on a Sigma Five digital computer. Total computation time for the two checks was slightly under one minute, or slightly under half a minute for each.

The biggest computational problems will occur in Block 2, Subblock 1 where m is about 40 and r is 6, and in Block 12, Subblock 3 where m is about 200 and r is 4. The first of these requires about 40^5, or 2^5 times as many terms as the check run. Since $2^5=32$ the probabilities for this cell can be calculated in about 15 minutes on a computer comparable to a Sigma Five. The second requires about 200^3, or about twice as many terms as the check run, and should only take about one minute. All required probabilities can be calculated in an hour or two.

The combinatorial formulae for p_1 through p_6 are:

$$p_1 = n_x/n \tag{1'}$$

$$p_2 = p_1 C_2 \tag{2'}$$

$$p_3 = p_1 \left[C_3 - \binom{m-2}{1} C_2 \right] \tag{3'}$$

$$p_4 = p_1 \left[C_4 - \binom{m-3}{1} C_3 + \binom{m-2}{2} C_2 \right] \tag{4'}$$

$$p_5 = p_1 \left[C_5 - \binom{m-4}{1} C_4 + \binom{m-3}{2} C_3 - \binom{m-2}{3} C_2 \right] \tag{5'}$$

$$p_6 = p_1 \left[C_6 - \binom{m-5}{1} C_5 + \binom{m-4}{2} C_4 - \binom{m-3}{3} C_3 + \binom{m-2}{4} C_2 \right] \tag{6'}$$

where

$$\binom{a}{b} = \frac{a!}{b!(a-b)!}$$

and

$$C_{k+1} = \sum_{\{i_1 \ldots i_k\}}^{*} \frac{n_{i_1} + \ldots + n_{i_k}}{n - n_{i_1} - \ldots - n_{i_k}} , \tag{7}$$

the sum being over all $\binom{m-1}{k}$ combinations of indices $i_1, \ldots, i_k$ taken from the first m integers but excluding x.

4. ESTIMATION

Notation. Let some characteristic of interest, C, be given. For example, C might be

C = a male between 40 and 50 years of age going to a hospital emergency unit to obtain treatment for an accidental injury associated with a power lawn mower.

It is assumed that each ERV included in the sample can be classified as either having, or not having, the characteristic C. Let

Y = total number of ERV's in the U.S. over a one year period which have the characteristic C.

Y is the quantity to be estimated. Let

h=1,...,13 denote Block subscript (Table 1);

i=1,2,3 denote Subblock subscript (Table 1);

$j=1,\ldots,r_{hi}$ denote sampled hospital within the $(hi)^{th}$ cell of Table 1, where

r_{hi} = number of hospitals sampled in the $(hi)^{th}$ cell (Table 2);

n_{hij} = number of ERV's within the j^{th} hospital of the $(hi)^{th}$ cell for the one year period in question;

w_{hij} = reciprocal of the probability, $(p_{hi} \ldots p_{r_{hi}})\, s/S$, that the j^{th} hospital of the $(hi)^{th}$ cell will be included in the sample;

q_{hij} = reciprocal $(1/p_1)$ of the probability, p_1, that this hospital was selected first for the $(hi)^{th}$ cell in Stage II.

Finally, let

$$Y_{hijk} = \begin{cases} 1, & \text{if the } k^{th} \text{ ERV for the } j^{th} \text{ hospital of the } (hi)^{th} \text{ cell has the characteristic C.} \\ 0, & \text{if not.} \end{cases}$$

The Estimator Equations. The number of ERV's having the characteristic C for the j^{th} hospital of the $(hi)^{th}$ cell is

$$\hat{Y}_{hij} = \sum_{K=1}^{n_{hij}} Y_{hijK} ,$$

and the estimated number of ERV's with C for the $(hi)^{th}$ cell is a weighted sum of these across the r_{hi} hospitals sampled in the cell, the weights being the reciprocals, w_{hij}:

$$\hat{Y}_{hi} = \sum_{j=1}^{r_{hi}} w_{hij} \hat{Y}_{hij} .$$

Finally, the estimate $\hat{Y}$ of Y is the sum of the $\hat{Y}_{hi}$ over all cells:

$$\hat{Y} = \sum_{h=1}^{13} \sum_{i=1}^{3} \hat{Y}_{hi} .$$

Combining the last two equations, the estimate $\hat{Y}$ of the total number of ERV's in the U.S. over a one year period which have the characteristic C is given by

$$\hat{Y} = \sum_{h=1}^{13} \sum_{i=1}^{3} \sum_{j=1}^{r_{hi}} w_{hij} \hat{Y}_{hij} . \qquad (8)$$

This is the equation for actual calculation of $\hat{Y}$. It is calculated from 2 sets of 119 numbers each. The first set, w_{hij}, consists of the reciprocals of the probabilities of selection for the hospitals. The second set, $\hat{Y}_{hij}$, consists of the total counts of ERV's having the characteristic C, by hospital.

Variance Estimation. Sampling of hospitals within a given cell is independent of sampling within other cells, so the variance of $\hat{Y}$ is the sum of the variances of the $\hat{Y}_{hi}$ across all 39 cells. The variance of $\hat{Y}_{hi}$ depends only on between hospital variation within the $(hi)^{th}$ cell. There is no within-hospital variation because sampling is complete within hospitals.

5. APPENDIX

Derivation of Equations (1') through (6'). For what follows let

$$N_i = n - n_i\,, \quad N_{ij} = n - n_i - n_j\,, \text{ etc.,}$$
$$n_{ij} = n_i + n_j\,, \quad n_{ijk} = n_i + n_j + n_k\,, \text{ etc.,} \qquad \text{(A1)}$$

and let

$$T_2 = \sum_i^{*} \frac{n_i}{N_i}\,, \quad T_3 = \sum_{ij}^{*} \frac{n_i n_j}{N_i N_{ij}}\,, \quad T_4 = \sum_{ijk}^{*} \frac{n_i n_j n_k}{N_i N_{ij} N_{ijk}}\,, \text{ etc.,} \qquad \text{(A2)}$$

where the sums are over all permutations of the indices excluding x.

Then from equations (A1), and (1) through (6), $p_1 = n_x/n$ and

$$p_2 = \frac{n_x}{n} T_2\,, \quad p_3 = \frac{n_x}{n} T_3\,, \quad p_4 = \frac{n_x}{n} T_4\,, \text{ etc..} \qquad \text{(A3)}$$

Next define

$$S_2 = \sum_i^{*} \frac{n_i}{N_i}\,, \quad S_3 = \sum_{ij}^{*} \frac{n_i}{N_{ij}}\,, \quad S_4 = \sum_{ijk}^{*} \frac{n_i}{N_{ijk}}\,, \text{ etc.,} \qquad \text{(A4)}$$

where the sums here are also over all permutations of the indices excluding x. Finally, define

$$C_2 = \sum_i^{o} \frac{n_i}{N_i}\,, \quad C_3 = \sum_{ij}^{o} \frac{n_{ij}}{N_{ij}}\,, \quad C_4 = \sum_{ijk}^{o} \frac{n_{ijk}}{N_{ijk}}\,, \text{ etc.,} \qquad \text{(A5)}$$

where now the sums are over all combinations of the indices excluding x, as denoted by the circles above the summation signs. Denote by P_b^a the number of permutations of a things taken b at a time:

$$P_b^a = a!/(a-b)!\,.$$

The derivation of (1') - (6') consists of showing that

$$\begin{aligned}
&0!\, T_2 = S_2 \\
&1!\, T_3 = S_3 - P_1^{m-2} S_2 \\
&2!\, T_4 = S_4 - 2P_1^{m-3} S_3 + P_2^{m-2} S_2 \\
&3!\, T_5 = S_5 - 3P_1^{m-4} S_4 + 3P_2^{m-3} S_3 - P_3^{m-2} S_2 \\
&4!\, T_6 = S_6 - 4P_1^{m-5} S_5 + 6P_2^{m-4} S_4 - 4P_3^{m-3} S_3 + P_4^{m-2} S_2
\end{aligned} \qquad \text{(A6)}$$

and showing that

$$S_K = (K-2)!\, C_K, \quad K = 2, \ldots, 6. \tag{A7}$$

Substituting (A7) in equations (A6), multiplying each side of the resulting equations by n_x/n, and using (A3) yields the desired equations (1') through (6').

The derivation of equations (A6) will be limited to showing that $T_3 = S_3 - P_1^{m-2} S_2$. The others are derived similarly. Thus, using the fact that $\frac{a}{b(b-a)} = \frac{1}{b-a} - \frac{1}{b}$, we get

$$T_3 = \sum_{ij}^{*} \frac{n_i n_j}{N_i N_j} = \sum_{ij}^{*} n_i \left(\frac{1}{N_i - n_i} - \frac{1}{N_i} \right) = \sum_{ij}^{*} \frac{n_i}{N_j} - \sum_{ij}^{*} \frac{n_i}{N_i}$$

$$= S_2 - (m-2) \sum_{i}^{*} \frac{n_i}{N_i} = S_3 - (m-2) S_2 = S_3 - P_1^{m-2} S_2.$$

To show (A7) with K=4, for example, note that by symmetry

$$S_4 = \sum_{ijk}^{*} \frac{n_i}{N_{ijk}} = \sum_{ijk}^{*} \frac{n_j}{N_{ijk}} = \sum_{ijk}^{*} \frac{n_k}{N_{ijk}}$$

so that

$$3 S_4 = \sum_{ijk}^{*} \frac{n_{ijk}}{N_{ijk}},$$

and the sum is over all permutations of i,j,k excluding x. But for fixed i,j,k the quantity n_{ijk}/N_{ijk} is invariant under the 3 permutations of i,j,k. Hence

$$3 S_4 = \sum_{ijk}^{*} \frac{n_{ijk}}{N_{ijk}} = 3! \sum_{ijk}^{o} \frac{n_{ijk}}{N_{ijk}} = 3!\, C_4,$$

so that $S_4 = (4-2)!\, C_4$. The general case is similarly shown. This completes the derivation of equations (1') through (6').

SUMMARY OF SEVERITY INDEX

Category 7 = Category 6's who are hospitalized and deaths = Severity Value of 2516

Diagnosis	Severity Category 6 Severity Value = 340	Severity Category 5 Severity Value = 81	Severity Category 4 Severity Value = 31	Severity Category 3 Severity Value = 17	Severity Category 2 Severity Value = 12	Severity Category 1 Severity Value = 10
Amputation	Any part of body					
Avulsion	25% of body +	head, eye, upper trunk	lower trunk	leg, arm, hand, foot, finger, toe	mouth, ear	
Burns	25% of body + or eye	all single body parts except finger,toe,ear			ear, finger, toe	
Cell Damage	25% of body +	head, face, eye, upper or lower trunk		leg, arm, hand, foot, finger, toe		
Concussion	25% of body +	head				
Contusion or Abrasion	25% of body +			head, upper trunk	ear, mouth, neck, eye, lower trunk	arm, leg, hand, foot, finger, toe
Crushing	head, arm, leg, trunk, foot, hand		finger, toe			
Dislocation	25% of body +	head, upper trunk	lower trunk, eye		arm, leg, hand, foot, finger, toe	
Foreign Body	25% of body +	head, upper trunk	lower trunk	mouth	arm, leg, hand, foot, finger, toe, eye	
Fracture	25% of body +	head, neck, upper and lower trunk	eye	arm, leg, hand, foot, finger, toe, mouth		
Hematoma	25% of body +	head, upper trunk	eye, lower trunk	arm, leg, hand, foot	finger, toe, ear, mouth, neck	
Internal Organ Injury	25% of body +	head, neck, upper or lower trunk	mouth, eye			
Laceration	25% of body +		head, eye, upper or lower trunk		arm, leg, hand, foot, finger, toe, ear	
Nerve Damage	25% of body +	all other body parts				
Puncture	25% of body +	head, face, upper trunk	eye or lower trunk		arm, leg, hand, foot, finger, toe, mouth	
Strain or Sprain	25% of body +			neck, upper trunk	lower trunk, eye	arm, leg, hand, foot, finger, toe, ear
	anoxia, electric shock, submersion	ingested or aspirated foreign object				
Dermatitis			25% of body +		head, face, eye, upper and lower trunk	arm, leg, hand, foot, finger, toe, ear

Appendix II

NEISS Injury Matrix