

ESTIMATING THE NUMBER OF HOMELESS IN CHICAGO

Peter H. Rossi, Social and Demographic Research Institute, University of Massachusetts, Amherst, MA 01003.

I: Introduction:

In the classical formulations of how best to apply social research to social policy and program issues, one of the recommended earliest steps is to develop intelligence about the problem in question, including forming adequate conceptual and empirical definitions, and assembling or gathering sufficient empirical data to provide a description of the problem in terms of its size, social location, and spatial distribution of the social problem in question.

In order to decide that homelessness in the United States is a significant social problem, it may be enough to know that there are some homeless persons in our society, that the condition of homelessness lies below the threshold of minimum decent living conditions, and to be quite certain that the number of homeless is increasing. However, in order to devise effective programs and policies and to allocate appropriate amounts and kinds of resources to operate those programs, it is essential to know with some confidence what is the total number of homeless and by what amount that number is changing. Before we can begin to design adequate social policies and programs addressing the needs of the homeless chronically mentally ill, it is clearly necessary to have valid understanding of the size, distribution, and composition of that population. Although exquisite precision in these numbers is probably unnecessary, magnitude estimation is essential, especially since there is so much uncertainty concerning magnitudes.

Although almost any non-zero number of homeless persons defines a disgraceful condition, it does make a considerable difference to social policy and program concerns whether the numbers are 350,000 (as HUD estimates of a few years ago suggest) or 1.5 millions or higher (as the National Coalition for the Homeless suggests). Knowledge about the distribution and social characteristics of the homeless is also crucial for policy reasons. It does make a difference whether the sex ratio is 100 or 300, what proportion suffer from substance abuse, whether homeless are composed primarily of long term unemployed, etc. Information on the compositional and distributional characteristics is clearly necessary for the design of appropriate sets of treatment, how best to deliver treatments, and so on.

With respect to the problem of homelessness, the classical formulation of applied social research has simply not been possible to implement. Awareness of and concern for the problem has far outstripped our knowledge about the dimensions of the problem. As is often the case, federal, state and local public officials moved quickly to put into place policies and programs designed to ameliorate the lot of the homeless without waiting for an adequate knowledge base to be developed. This is not atypical for social problems: Often enough policies and programs are in place long before we know what may be the details of the problem. But there are special technical difficulties in developing empirically grounded knowledge about homelessness that have slowed considerably the development of such empirical knowledge, about which we will say more later in this paper. Of course, this does not mean that there are no ideas of the size and composition of the homeless population that lie behind what is being done: It does mean that those conceptions are based on guesstimates and conjectures, bolstered by dramatic anecdotal accounts and fervent personal testimony.

This paper provides an account of a research endeavor that attempted to meet some of the obstacles to the empirical study of homelessness, the Chicago Homeless Study, funded by grants from the Robert Wood Johnson Foundation, the

Pew Memorial Trust and the Illinois Department of Public Aid. Conducted in 1985 and 1986, we believe that this study is the first to adapt conventional sample survey methods to the study of a population that is uniquely unsuited to study by conventional sample survey methods. The goal of the Chicago Homeless Study was to devise new sample survey approaches and to demonstrate their utility.

II: Obstacles to Research on the Size, Composition and Distribution of the Homeless.

Compared to the usual survey operation, a survey of homeless persons presents a number of special problems that are obstacles to be overcome in survey design.

A. Definitional Issues:

A primary obstacle is the absence of a widely accepted definition of homelessness. Indeed the range of definitions is very wide, running the gamut from highly inclusive definitions that cover all persons who are inadequately housed--including doubled up households, persons living in (by some standards) poor housing, persons temporarily housed in hospitals or other institutionalized contexts, as well as those who do not rent or own a conventional dwelling unit. More restrictive definitions focus primarily on those who do not rent or own conventional dwelling units.

Definitional issues are not simply technical ones: Involved are value issues centering around what should be considered the floor of decent housing below which no one should be allowed to sink. Clearly, more inclusive definitions enlarge the size (and also change the composition) of the homeless. They also enlarge and considerably complicate the task for social science researchers. More inclusive definitions also place very fuzzy boundaries around homelessness allowing guesstimates to range more widely.

In practice, operational definitions tend to be less inclusive, focusing primarily on what my colleagues and I have called the "literal homeless", as opposed to "precariously domiciled". This occurs because the more inclusive definitions are extremely difficult to implement, except at prohibitive costs.

B. The Primary Obstacle: Developing a sampling frame for the homeless:

Social science research has made considerable strides in the last half century. Censuses are considerably more accurate and precise; sample surveys have been developed to the point where much of our information about the size, composition and distribution of modern population are known with a precision that would have been unthinkable half a century ago. However, conventional social research methodologies are largely useless in the study of the homeless since most of these methods are based on the assumption that, for all practical sampling purposes, we can assume that every person or household has an address at which that person may be reached either by an interviewer going to that address, through a telephone number or by mail. This critical assumption simply does not hold for the bulk of homeless persons by definition. This means that sampling frames for surveys of homeless have to be quite different than conventional sampling frames for the domiciled population.

C. Statistical Rarity:

Homelessness may be, as I believe it to be, a national disgrace. Nevertheless, under even the most inclusive definitions of homelessness, it is still a very rare condition that affects at most -- under the most inclusive definition -- 1.5% of the adult population (and most likely) affects less than 1/10 of 1% of the adult population. The study of rare populations is not impossible: but it is expensive. For example,

if we undertook a random sample of any urban area, in order to obtain a sample of homeless persons, we would have to approach anywhere from 70 to 500 adult persons in order to encounter one homeless persons.

D. Identification

Whatever definition of homelessness you may adopt, there remains the problem of how to apply it in specific instances. Homelessness is simply not directly observable so that the application of most definitions require obtaining information directly from individuals. The problem of identification is one that plagues the use of client data in the study of homelessness. For example, none of the R W Johnson medical clinics for the homeless actually attempt to determine whether a person who presents himself or herself for medical services is homeless. (This is most likely a sensible strategy.) Our best bet is that some portion--size unknown -- of their clients are not homeless by our definition.

E. Transiency and Turbulence

An important characteristic of the homeless population is that its composition changes constantly. Under any definition the line between the homeless and the domiciled is often crossed, and in both directions. Some of the important distinctions that one needs to be aware of are as follows:

a. One time Momentary (or very short term) homeless: This group consist of persons who are homeless for less than a week and only once or twice over say a few years. For example, runaway or throwaway young people who are usually rejoined to their families within a few days.

b. Part-time or Periodic Short term homeless: Persons who are homeless repeatedly in a somewhat regular pattern, say the last few days before their SSI checks arrive.

c. Transitional homeless: persons homeless because they are in transition between one state and another and who have the resources or potential to connect with the conventional housing market-- e.g. young mothers and their children leaving one household arrangement, households burned out or evicted, persons migrating into a new community, etc.

d. Long term homeless: Persons homeless and without the potential or actually in hand resources to connect with the conventional housing market.

The heterogeneous distribution of time homeless means that it would be very useful to compute a variety of prevalence and incidence estimates, as measures of the size of the homeless population.

E. Heterogeneity

As we have come to learn from studies of the homeless, they are quite heterogeneous in age, gender, degree of disability and in other ways. This heterogeneity undoubtedly reflects corresponding heterogeneity in the sources of homelessness, as well as corresponding differences in the kinds of treatments that would be appropriate. Heterogeneity also applies to their micro-location: some portion of the homeless can be located through the shelters provided for homeless persons but some other portion (of unknown size) cannot be reached in that fashion. The heterogeneity of the homeless makes it important to have sample sizes that are large enough both to identify and to study in detail critical subgroups within that population.

F. Geographical Concentration

The homeless are not distributed uniformly throughout a community. In part, this is a plus in studying them but in part it is an obstacle. The concentration of homeless persons occurs because the institutions that serve the needs of the homeless are also concentrated, but in turn that institutional concentration reflects the concentration of the homeless. From a sampling point of view, this means that optimal sampling requires some prior knowledge of the spatial distribution of the homeless.

F. Communications Problems:

In addition to the usual problems encountered in attempting to survey poorly educated persons, the homeless are more difficult to interview because of the prevalence of disabilities -- including substance abuse and mental illness-- that make communications difficult. This is not a counsel of despair -- among the close to 800 homeless persons we tried to interview in Chicago, in less than half a dozen instances was it impossible to obtain any information.

III: Alternative Approaches:

There are several approaches that have been used in studying the homeless, varying widely in validity

A. Key Person Surveys: Essentially asking persons with some knowledge and acquaintance with homelessness to provide their estimates of the size and composition of this group. This is the approach used both by HUD, and the National Coalition for the homeless, leading to extremely disparate estimates.

The main problem with key persons surveys is that they are of unknown validity -- probably better than no information at all but how much better no one knows.

B. Partial Counts:

Surveys undertaken of some portion of the homeless population that can be easily identified: surveys of persons in shelters, on lines of soup kitchens, at known gathering places of homeless persons. e.g. the 1980 Census' so-called "casual count". These surveys are better than sheer guessing but again of unknown validity for estimating purposes because the extent of bias in partial counting is unknown.

The main problem with partial counts is that no one knows what is the proportion of the total homeless population that is being omitted. Nor does it make any sense to assume that there is some constant ratio between components, as was the case for HUD's estimates, premised on a constant ratio of shelter inhabitants to total homeless.

C. Windshield street surveys

These are counts of homeless persons undertaken by identifying homeless persons by sight while canvassing the streets and other open places in a city. This was the procedure used in surveys of Washington DC, and Boston Mass. (The Washington DC study actually included more than a windshield survey.) There are some very attractive features of windshield surveys -- they are relatively inexpensive, bypass the communications problems, but are biased to the extent that homeless persons are not on the streets or in public where they can be viewed through windshields. Furthermore, it is not at all clear that homeless persons can be identified easily by their appearance or that all domiciled persons do not resemble homeless persons in appearance.

D. Adaptations of Area probability designs

For the time being, the only study that has adapted area probability designs for the study of the homeless is the Chicago Homeless Survey conducted by my colleagues and myself, using NORC as the subcontracting data collection agency. The remainder of this paper describes this survey in detail.

IV: The Chicago Homeless Study

This section contains a brief description of the methodology of the Chicago Homeless and Study and points out the ways in which the various obstacle I listed earlier were met.

A. Definition employed: We defined our target population as residents of the City of Chicago who did not rent or own (or has rights of customary access to a conventional dwelling unit -- an apartment, a house, a mobile home, or a room or rooms rented in a hotel, SRO, rooming house, mobile home. The intent of our definition was to cover the "literal homeless", a decision that subjected the study to a great deal of criticism. Note that this definition does not include among the homeless those who are doubled up, or living in inadequate quarters.

The adoption of this restricted definition does not, of course, solve the definitional issues. At the most, it simply clarifies where this study should fit in the full array of all possible studies using different definitions.

B. Sampling Frame:

In relation to a conventional sampling frame, the sampling frame used in this study is like a photographic negative of the conventional sample survey, being a sample of non-dwelling units, as opposed to the usual definition of dwelling units. Our survey defined two subsamples:

1. A probability sample of shelters provided primarily for homeless persons, chosen with probabilities proportionate to shelter size. Within the sample of shelters, a systematic sample of shelter residents was taken, amounting to one in three sampling ratio. We enumerated the total universe of shelters, using experts as our guide. NORC interviewers collected rosters of shelter residents and systematically sampled from those rosters.

2. A unbiased probability sample of Chicago blocks on which the non-dwelling unit areas and places on the blocks would be of primary interest. Interviewers would search the non-du places on each of the sampled blocks, interviewing each person encountered to determine whether or not homeless. If a person encountered was determined to be homeless, she or he was interviewed. Non-du places on each block were defined as any place to which access could be obtained without encountering a locked door. Non-du places that were routinely searched included streets and sidewalks, alleys, doorways, hallways, garages, abandoned buildings, open basements or roofs. In addition, interviewers were instructed to examine parked cars, trucks, boxcars, etc.

This implies that the operational definition of homelessness consists of persons who are shelter residents or on the streets or in public places who did not own or rent some conventional dwelling unit.

C. Statistical rarity, Identification and Geographic Concentration:

To make the sample of blocks more efficient and to minimize the screening problem posed by the identification problem, we did the following:

1. Stratified the 19,000 blocks of Chicago by obtaining expert guesses concerning the blocks of Chicago. We obtained the cooperation of the Police Department to have each of the blocks of Chicago stratified into three categories: High, medium and low probabilities of finding homeless persons on each block. The resulting stratification was reviewed by homeless experts. This allowed us to draw random samples of blocks within each of these three strata. The prior stratification was not perfect, nor need it be. There is a nice characteristic of this sort of prior stratification. If it is good, it can increase efficiency. However, if it is poor it does not bias the sample, although the resulting sample may be inefficient.

Because we anticipated that the size of the Chicago homeless sample would vary seasonally, reinforced by our experiences in September and October of 1985 when we conducted our first survey, we planned and carried through an additional survey in the dead of the winter in February and March 1986. As we will show later on, our first survey was under-designed, a flaw that our second survey remedied. These two surveys are designated as Phase I and Phase II respectively on the tables in the handout.

Details of the sampling designs for the two surveys and for both the shelter and street subsamples is given in Tables 1 and 2. Tables 3 and 4 present the field experiences of the 2 subsamples, indicating, I believe a fairly good completion rate for both subsamples in both of the surveys.

2. Reducing the identification problem: Conduct the survey at times when domiciled or homed persons were least likely to be encountered. The interviewing took place between the hours of midnight and 6 AM.

3. Chicago (or any other city) is not a safe place in the dead hours of the night. In order to protect our interviewers, we hired off duty policemen as escorts, who accompanied the NORC interviewers as they made their rounds on the streets and in the shelters.

D. Transiency:

Obtain information on how long homeless so that the time homeless distribution could be estimated without making too many heroic assumptions. Using this distribution we were able to estimate average nightly prevalence of homeless for two periods, Fall 1985 and Winter 1986, annual incidence and annual prevalence

E. Communication:

Pay respondents \$5 per interview. Although payment did not reduce the incoherence of some of the subjects, it did assure that cooperation was very good. The end result was an unbiased sample of the literal homeless of Chicago.

Two surveys were conducted, using much the same approach, one in the Fall of 1985 and the other in the Winter of 1986. The combined surveys provided data on the characteristics of the homeless of Chicago, details of which have been published elsewhere.

The two surveys also provided unbiased estimates of the sizes of the literal homeless populations of Chicago, as shown in the Table 5 of the handout. Note the much larger standard errors for the first sample in which we did not have a large enough sample of blocks to counter the heterogeneity of block estimates. The Phase II block sample is considerably larger. These estimates have not been greeted with joy by the advocates of the homeless in Chicago. The prevailing guesstimates had pegged the size of the Chicago Homeless as being between 15,000 and 25,000. The estimates shown in Table 5 are many magnitudes smaller than the lowest prevailing ones.

IV: Evaluating the Chicago Homeless Survey Approach

There are both limitations and advantages to the approach we took in Chicago, as follows:

1. Cost: The Chicago Homeless Study was quite expensive, the total bill for both surveys being close to \$250,000. Of course, a lot of those costs were one time development costs and some were caused by errors that we now know how to avoid. Still I doubt whether we would replicate the study in Chicago for less than \$100,000.

2. Definitional limitations: The survey self consciously limited itself to the literal homeless. Although in principle it is possible to expand the definition to the marginally or precariously homed, to do so would mean to mount an even more extensive survey operation in which existing domiciled households would have to be screened to identify the precariously homed.

3. Response Validity: The data obtained in the survey consist of self reports given by the homeless. Given the level of disability and substance abuse prevalent in this population, the issue of response validity can be raised appropriately. We have no way to assess the importance of this issue.

The advantages of the approach we used are also formidable.

1. Credibility: Because the approach is based on sound sampling rationale, its results are credible.

2. Generalizability: The approach is not site-specific. The operations we undertook in Chicago can be replicated in any American political jurisdiction. Indeed, like other area probability approaches it can be generalized to the nation as a whole and is capable thereby of producing credible national estimates of the homeless population of the United States.

Table 1

**Chicago Shelter Universe and Shelter Samples
Phase I and Phase II^a**

A. Shelter Universe and Samples: Phase I and II	Phase I	Phase II
1. Eligible Shelters in universe -----	28	45
2. Universe bed capacities -----	1573	2001
3. Shelters drawn in sample -----	22	27

B. Details of Phase II Shelter Sample

Shelters were drawn with probabilities proportionate to size, with residents sampled disproportionately within shelters to form a self-weighting sample. Sampling ratios for the Phase II sample were as follows:

Shelter Size Classification	Number in Universe	Number in Sample	Occupant Sampling Ratio
Large Shelters (37+ Beds)	17	17	.25
Medium Shelters (18 to 33 beds)	12	6	.50
Small Shelters (less than 18 beds)	16	4	1.00

^a Phase I survey conducted during the last week of September and the first week of October 1985. Phase II survey operations spanned the last week and first week respectively of February and March 1986.

Table 2

Street Survey Sample Design

I: Stratified Random Block Sample: Phase I and Phase II^a

Census Block Classification and Sample Sizes

Prior Density Classification ^b	Universe Number of Blocks	Phase I Sample Size	Phase II Sample size
High Density	295	49	49
Medium Density	806	49	49
Low Density ^b	18308	70	147
TOTAL	19409	168	245

^a Phase I and Phase II samples were drawn independently.

^b Prior classification accomplished with the help of community relations officers of the Chicago Police Department and modified with the help of other knowledgeable persons.

^c Low density blocks were sampled in clusters of five blocks in Phase I and of three blocks in Phase II.

Table 3

Shelter Survey Implementation Experiences

	Phase I	Phase II
A. Shelter Universe		
1. Eligible Shelters in universe -----	28	45
2. Shelters drawn in sample -----	22	27
3. Sampled Shelters agreeing to participate -----	21	23
Shelter Completion Rate -----	(95.5%)	(85.2%)
B. Shelter Resident Sample		
1. Eligible residents in sampled shelters-----	934	1183
2. Eligible residents selected in sample-----	320	317
3. Sampled residents interviewed-----	265	248
Completion Rate-----	(82.8%)	(78.2%)
4. Uninterviewed Sample Shelter Residents		
a. Unavailable ^a for interview -----	41	19
	(12.8%)	(6.0%)
b. Refused interview-----	14	49
Refusal Rate----- (4.4%)		(15.4%)
c. Interview breakoff before completion-----	0	1

^a "Unavailable" means that person was not present in the shelter at the time of the survey (e.g. at work, temporarily absent, etc.) and hence was not asked for interview. Callbacks were attempted on all unavailables: The cases shown here were those with whom callbacks were not effective in retrieving the interviews.

Table 4

Street Surveys Implementation Experiences

	Phase I	Phase II
A. <u>Block Universe And Sample</u>		
1. Blocks in universe -----	19409	19409
2. Blocks selected in sample-----	168	244 ^a
B. <u>Street Search Experience</u>		
1. Persons encountered ^b and approached for screening	318	289
2. Persons encountered <u>and</u> screened -----	232	238
Screen Completion Rate----- (73%)		(82%)
3. Persons who refused screening interview -----	80	37
Screen Refusal Rate -----	(25%)	(13%)
4. Screen interview breakoff -----	4	5
5. Person encountered unable to be screened -----	2	9
C. <u>Main Questionnaire Experiences</u>		
1. Persons screened and eligible for main interview	23	30
2. Completed main interview -----	22	28
Completion Rate -----	(96%)	(93%)
3. Refused main interview -----	1	0
Interview Refusal Rate -----	(4%)	(0%)
4. Interview breakoff before completion-----	0	2

^a Note that 245 blocks were selected for the block sample. However, one of the blocks selected "did not exist" and hence was not searched.

^b An "encounter" consists of any person present on a block or in any of the public access places on a block who was walking, sitting, standing, lying down, sitting in a parked car or truck, or riding a bicycle.

Table 5

**Phase I and Phase II Point Prevalence Estimates
of the Chicago Homeless Population**

Point Prevalence Estimates = Average Daily Number of Persons Homeless
During the Relevant Survey Period

A. Phase I Estimates: (September 22 - October 4, 1985)

Survey Component	Estimate	Standard Error
Shelter residents -----	961	± 13
On streets or in public places -----	1383	± 735
<u>Total Homeless</u> -----	2344	± 735
Range ± 1 Stand. Error = 1609 - 3079		

B. Phase II Estimates: (February 22 - March 7, 1986)

Survey Component	Estimate	Standard Error
Shelter residents -----	1492	± 55
On streets or in public places -----	528	± 269
<u>Total Homeless</u> -----	2020	± 275
Range ± 1 Stand. Error = 1745 - 2,295		