

## USE OF CENSUS MATCHING FOR STUDY OF PSYCHIATRIC ADMISSION RATES

Earl S. Pollack, National Institute of Mental Health

Studies or analyses designed to measure the rate of occurrence of a particular event in specific population groups are extremely common. In most of those concerned with illness or mortality, the numerators are obtained from interviews, vital records, or hospital or agency case records and are related to published population data. The assumptions implicit in such a procedure are : (1) that each individual counted in the numerator has been enumerated in the population and (2) that each individual is classified identically in both numerator and population denominator with respect to the characteristics under study.

An alternative procedure involves identifying the individuals to whom the event of interest has occurred and locating for each of these persons the Census document used for tabulating population data. If this procedure is successful in locating the Census records for all of the persons in the study, both of the above assumptions will be fulfilled. It is the purpose of this paper (1) to describe a study using this procedure, (2) to present data indicating the relative success of the Census matching procedure for various groups and (3) to discuss the implications of failure to find matching Census schedules for the analysis of rates.

The Psychiatric Admission Rate Study

In 1960, a study was begun to determine the rates at which persons come under psychiatric care in specific population groups defined primarily by a number of socioeconomic and family relationship variables. Data of this type were needed to help plan for the development of programs of psychiatric care and to provide a set of hypotheses for further study into the etiology of mental disorders. Further impetus to the formulation of the study was provided by the impending availability of a large volume of population data to be prepared from the 1960 Census.

Two states, Maryland and Louisiana, were selected as the locale for the study on the basis of their extensive programs for central reporting of data on persons coming under care in inpatient and outpatient psychiatric facilities and because of the interest in the study expressed by key persons in the agencies responsible for the mental health programs in these states. With the cooperation of the Louisiana State Department of Hospitals and the Maryland Departments of Mental Hygiene and of Health, the Office of Biometry of the National Institute of Mental Health collected basic identifying information on each person admitted to the public and private inpatient and outpatient psychiatric facilities in the two states during the year following the Census. This information was given to the Bureau of the Census, where 1960 Census

schedules for these individuals were located, and detailed tabulations of the demographic, socioeconomic and family relationship characteristics of these patients were tabulated.

Matching Method

Since the major focus of this paper is on the results of the matching procedure, the procedure itself will be described only briefly. For each person admitted to a psychiatric facility in the two states during the study period, a transcription sheet was submitted to the Bureau of the Census containing the following information: name, sex, color, date of birth, psychiatric diagnosis, facility to which admitted, history of previous admissions, residence as of time of admission and as of April 1, 1960, and name of head of household on that date.

It should be emphasized that the matching was carried out by hand at the Bureau of the Census, not by computer. The Census schedules are filed by enumeration district (ED), a small geographic subdivision assigned to a single census enumerator and consisting of an average of about 250 housing units. Therefore, the success of the matching operation depended heavily on the accuracy of the address for each person admitted to a psychiatric facility. The transcription sheets were sent first to the geography unit where the appropriate ED number was assigned to each address. If an address was given in rather vague terms, it could have been assigned with equal justification to more than one ED. Therefore, on the transcription sheet, space was allotted for the assignment of a maximum of seven ED's for a given address.

The transcription sheets were next sent to a processing unit where an attempt was made to locate the Census schedule corresponding to each individual in the study. The schedules are filed in books according to ED. The search was carried out in two stages: (1) finding the page in an ED book which contained the same address as that given for the patient on the transcription sheet and (2) identifying the patient on that page. A set of rules was provided for each of these stages. On transcription sheets where several possible ED's were indicated, each was searched in turn until the address was found. If the address could not be found in any of the ED books indicated and if a search for the patient's name in each of those books also proved unsuccessful, the patient was considered a "non-match." Further procedures were involved for those individuals who were included in the census 25 percent sample. Since this is not important for this presentation, the discussion which follows will pertain primarily to the census 100 percent data.

### Results of Matching

Transcription sheets containing basic information on 13,036 Louisiana patients and 14,450 Maryland patients were submitted to the Bureau of the Census for purposes of locating the corresponding 1960 Census schedules. Of these, matching schedules were found for 67 percent of the Louisiana patients and 64 percent of the Maryland patients. The fact that the matching procedures failed to locate census schedules for approximately one-third of the patients raises a serious question about the validity of admission rates based on matched cases only. An analysis of match rates according to specific characteristics and in relation to existing knowledge about completeness of census enumeration will help to place in perspective the effect of these match rates on the analysis of admission rates.

The extent of matching varied considerably from one category to another. Match rates according to specific characteristics are presented in Tables 1 through 7. The following are a few of the highlights:

1. Matching was most successful among those under 18 years of age; least successful among those 18 to 24; higher among males than females; higher among whites than non-whites.
2. Match rates were far lower for alcoholics than for any other diagnostic group.
3. For household heads and members of their immediate families, three-fourths of the matching schedules were found, whereas the match rates for other relatives were 63 percent in Louisiana and 57 percent in Maryland and for non-relatives only 40 percent and 49 percent, respectively.
4. Among married persons the match rates were 77 percent for Louisiana and 76 percent for Maryland.

These rates have been presented in some detail in the tables with the hope of illuminating reasons for failure to find matching Census schedules. Possible reasons for this failure are:

1. Inadequate or poorly defined addresses.
2. Differences between census and admission records in name and age.
3. Clerical errors.
4. Persons not enumerated in the census.

No specific studies have been conducted which would permit a classification of non-matched cases into these four categories. Comparison of the above results with those of other census matching studies reveals consistently lower match rates for the present study (1,2,3). The study most comparable to the present one was

that conducted by the University of Chicago in cooperation with the National Vital Statistics Division in which deaths during the four-month period following the 1960 Census were matched against Census schedules. Preliminary analysis indicates that the overall match rate obtained in that study was approximately 80 percent (4). Since the Census matching for this study was carried out immediately prior to that for the present study using the same clerical staff and almost all of the same procedures, one would expect the quality of the search to be comparable between the two studies.

It is likely, therefore, that differences in match rates between the two studies are due to differences in quality of the addresses used as a basis for matching, differences in the extent to which those under study had been enumerated in the population, or some combination of these two factors.

Unfortunately, based on information presently available, there is no way of determining the relative influence of each of these two factors. Evaluation of the 1960 Census is still being carried out, but some preliminary data providing estimates of the net census undercount according to sex, color and age are presented in Table 8 (5). These undercounts, particularly among the whites, may seem too low to have any appreciable effect on the match rates. It should be pointed out, however, that these are estimated average counts over the total population, and furthermore, that they pertain to the population of the United States rather than to those of the two states under study.

As a result of the studies evaluating the 1950 Census, the rates of omission from the census by household composition were estimated as follows: Head or wife, 2.0 percent; child of head, 1.7 percent; other relative of head, 4.1 percent; unrelated individuals, 8.2 percent (6). The investigators who carried out these studies stated that a sizeable proportion of persons in enumerated households who were omitted from the Census may be persons with no regular place of residence.

What, then, can be said about the extent of Census under-enumeration among patients included in the present study? Based on the data presented above the following statements can be made: (1) In categories in which match rates tend to be high (Table 1) census undercounts tend to be low (Table 8). (2) Table 4 indicates high match rates for household heads and immediate members of their families, lower rates for other relatives and very low match rates for unrelated individuals. As indicated above the rate of omission from the census was lowest among household heads and members of their families; next lowest among other relatives; and highest among unrelated individuals. (3) It seems reasonable to expect that Census enumeration among alcoholics and persons living alone would be more difficult than for the population in general. If this is indeed true

one would expect low match rates among these groups due to underenumeration alone.

#### Implications of Nonmatching and Underenumeration for Analysis of Admission Rates.

Suppose we wish to compare rates of admission to psychiatric facilities between two subgroups of the population, say, single and ever married. Ideally, we would divide the total number of single persons admitted to psychiatric facilities by the total number of single persons in the population and compare the result with a similar ratio for married persons. In this study two factors complicate this comparison: (1) the numerator of this ratio is incomplete due to the inability to find all of the Census schedules, and (2) the denominator is also understated because of underenumeration of the population in the census. To complicate this problem further, the extent to which Census schedules were found is unknown for many of the categories to be considered in the analysis and the proportion of underenumeration is unknown for every category of the population.

How, then, can we obtain a valid comparison of the admission rates when only incomplete counts of both numerator and denominator are available? How can we make use of the available data to approximate the results that we would obtain if complete counts of both numerator and denominator were available? These questions can be answered more readily if we consider first the following formulation of the problem.

Consider, again, a comparison of rates of admission to psychiatric facilities between two categories of the population, single and ever married.

Let  $Y_i$  = number of admissions in the  $i^{\text{th}}$  marital status category

( $i = 1, 2$  for single and married, respectively)

$P_i$  = total population in the  $i^{\text{th}}$  category

$y_i$  = number of matched admissions

$p_i$  = enumerated population

$m_i$  = proportion of admissions matched to Census schedules

$e_i$  = proportion of the population in the  $i^{\text{th}}$  category which was enumerated in the census.

$\hat{Y}_i = y_i + \text{an estimate of } (Y_i - y_i)$

$R_i = Y_i/P_i$

$r_i = y_i/p_i$

$r'_i = \hat{Y}_i/p_i$

Ideally, we would like to compute the "true" admission rate,  $Y_i/P_i$ , but complete data on neither the numerator nor the denominator are

available. There are two alternatives open to us:

- (1) compute  $Y_i/P_i$
- (2) compute  $\hat{Y}_i/p_i$  by adding an estimate of the number of nonmatched cases in the  $i^{\text{th}}$  category to the known number of matched cases.

But  $Y_i/P_i$  is a "good" estimate of  $Y_i/P_i$  only if  $m_i/e_i$  is close to unity, and  $\hat{Y}_i/p_i$  will, on the average, overestimate  $Y_i/P_i$ , because the expected value of  $\hat{Y}_i = Y_i$ , but  $p_i$  is almost always less than  $P_i$ .

In comparing rates between two marital status categories, however, the problem is not necessarily that of obtaining "good" estimates of  $R_i = Y_i/P_i$ , but rather to obtain "good" estimates of  $R_1 - R_2$ , if we are interested in the excess risk of admission among single persons or  $R_1/R_2$  if we are interested in the risk of admission among single persons relative to that among married.

For purposes of this presentation only the relative risk,  $R_1/R_2$ , will be considered. Using data on matched admissions and enumerated population, the relative risk can be written.

$$\frac{r_1}{r_2} = \frac{m_1 Y_1}{e_1 P_1} \cdot \frac{e_2 P_2}{m_2 Y_2} = \frac{Y_1}{P_1} \cdot \frac{P_2}{Y_2} \cdot \frac{m_1 e_2}{m_2 e_1} = \frac{R_1}{R_2} \cdot \frac{m_1 e_2}{m_2 e_1},$$

where  $R_1/R_2$  is the "true" relative risk. If the ratio  $m_i/e_i$  is relatively constant, i.e., if  $m_1 e_2 / m_2 e_1 = 1$ , the observed relative risk is approximately equal to the "true" relative risk.

If we add to the numerator the estimated number of nonmatched admissions, the resulting relative risk is

$$\frac{r'_1}{r'_2} = \frac{\hat{Y}_1}{e_1 P_1} \cdot \frac{e_2 P_2}{\hat{Y}_2} = \frac{Y_1}{P_1} \cdot \frac{P_2}{Y_2} \cdot \frac{e_2}{e_1}$$

If  $e_1 = e_2$ , the observed relative risk is equal to the true relative risk.

This formulation suggests that to simplify the interpretation of ratios of admission rates, two conditions must be fulfilled:

- (1)  $\frac{m_1 e_2}{m_2 e_1} = 1$  when using rates of the form  $\frac{Y_i}{P_i}$
- (2)  $\frac{e_2}{e_1} = 1$  when using rates of the form  $\frac{\hat{Y}_i}{P_i}$

When these conditions hold, ratios of rates provide consistent estimates of the "true" relative risks.

To obtain some idea of the extent to which these ratios might deviate from unity, examples of each of these two types of ratios were computed based on the data presented in Tables 1 and 8 and are shown in Table 9. It should be reiterated, however, that the data on census undercounts, presented in Table 8, are merely crude estimates for the entire United States.

It will be noted that the greatest deviation from unity for the ratio  $m_1 e_2 / m_2 e_1$ , occurred in the age group 65 and over in each state and resulted from low match rates among non-white males. These were not offset by the corresponding enumeration rate which, far from being low, was estimated as a 7.9 percent overcount. Aside from those categories, two-thirds of the ratios differed from unity by less than 10 percent. Similarly, more than two-thirds of the ratios  $e_2 / e_1$  deviated from unity by less than 10 percent.

This provides some indication that if similar ratios could be obtained, based entirely on Louisiana and Maryland data, the assumptions made in assessing relative risks of admission to psychiatric facilities will be fulfilled with relatively small error. Except for the meager pieces of evidence provided thus far by the Census post-enumeration surveys, the  $e_i$  are unknown. Match rates, on the other hand, can be estimated for some variables, but are unknown for others. Therefore, since only part of the information required to make a choice between the two alternatives is available, both rates,  $y_i / p_i$  and  $\hat{y}_i / p_i$ , will be computed for each category, where possible. If the results of a given comparison are consistent for the two sets of rates, they can be interpreted with greater confidence, perhaps, than results based on only one set of rates. If, on the other hand, conclusions differ between the two sets of rates, they will be viewed as inconclusive. In such an event, however, some interpretation will be made based on the knowledge available on match rates and enumeration rates for the categories involved.

The findings of this investigation could have rather far-reaching implications for studies in which numerators of rates are not obtained from matching Census schedules. Such numerators are essentially complete, but the corresponding denominators are incomplete due to underenumeration of the population. Marked variation in the extent of underenumeration among the categories being compared could result in substantial differences in rates due to differences in Census coverage alone. A careful comparison of the results obtained from the two sets of rates,  $y_i / p_i$  and  $\hat{y}_i / p_i$ , may shed some light on this problem. In addition, more detailed data on the extent of Census coverage in segments of the population defined by a number of variables would be extremely helpful.

## REFERENCES

1. Kaplan, D.L., Parkhurst, E. and Whelpton, P.K.: The comparability of reports on occupation from vital records and the 1950 Census. Vital Statistics - Special Reports, U.S. Dept. of H.E.W., Public Health Service, 53, June 1961, 1.
2. General Register Office: 1951 Census of England and Wales, General Report. London, H.M. Stationery Office, 1958.
3. Guralnick, L. and Nam, C.B.: Census - NOVS study of death certificates matched to census records. Milbank Mem. Fund. Quart., April 1959, 144.
4. Hauser, P.M. and Lauriat, P.: Record matching -- theory and practice. Proceedings of the Social Statistics Section, Am. Stat. Assn., 1963.
5. Taeuber, C. and Hansen, M.H.: A preliminary evaluation of the 1960 censuses of population and housing. Unpublished paper, Sept. 1963.
6. U.S. Dept. of Commerce, Bureau of the Census: The Post-enumeration Survey, 1950, Technical Paper No. 4. Washington, D.C., U.S. Bureau of the Census, 1960.

TABLE 1

Percent of Matching 1960 Census Schedules Found for  
Psychiatric Admissions, Louisiana and Maryland,  
by Age, Sex and Color

Age, Sex and Color	Louisiana		Maryland	
	Total persons admitted	Percent matched	Total persons admitted	Percent matched
All classes	13,036	66.8	14,450	63.9
Under 18 years	2,643	76.5	3,428	70.8
18-24	1,235	55.9	1,232	55.0
25-44	5,432	64.4	5,795	59.9
45-64	2,982	67.2	2,942	66.1
65 years and over	744	66.7	1,053	67.4
White males	5,258	66.3	6,167	64.3
Under 18	1,412	78.0	1,850	71.7
18-24	381	54.3	430	57.2
25-44	1,919	60.7	2,112	58.8
45-64	1,278	65.4	1,370	65.0
65 and over	268	66.0	405	65.2
White females	4,549	71.1	5,255	69.5
Under 18	765	76.5	955	72.3
18-24	456	59.0	456	56.8
25-44	2,013	70.8	2,235	68.7
45-64	1,032	72.7	1,083	73.0
65 and over	283	72.1	526	71.7
Non-white males	1,533	59.2	1,710	51.1
Under 18	241	76.3	393	65.1
18-24	183	53.6	183	47.5
25-44	674	56.5	782	45.4
45-64	331	58.3	281	49.8
65 and over	104	50.0	71	49.3
Non-white females	1,696	64.0	1,318	56.1
Under 18	225	67.6	230	67.0
18-24	215	54.0	163	52.8
25-44	826	64.0	666	51.1
45-64	341	66.3	208	60.1
65 and over	89	70.8	51	66.7

TABLE 2

Percent of Matching 1960 Census Schedules Found for  
Psychiatric Admissions, Louisiana and Maryland,  
by Selected Mental Disorder

	Louisiana		Maryland	
	Total persons admitted*	Percent matched	Total persons admitted*	Percent matched
All diagnoses	13,036	66.8	14,450	63.9
Alcoholism	1,255	54.7	1,612	50.9
Diseases of the Senium	510	62.5	755	63.7
Schizophrenia	3,382	62.4	3,322	58.7
Psychoneurotic reactions	2,449	74.2	2,280	70.2
Transient situational personality disorders	940	76.5	1,263	74.0
All other diagnoses	3,626	66.5	4,054	66.1
Undiagnosed	874	74.7	1,164	65.8

\*Persons admitted more than once were counted only once and if more than one diagnosis was given for a person, the first one was used for this table.

TABLE 3

Percent of Matching 1960 Census Schedules Found for  
Psychiatric Admissions, Louisiana and Maryland  
by Type of Psychiatric Facility\*

	Louisiana		Maryland	
	Total persons admitted	Percent matched	Total persons admitted	Percent matched
Public mental hospitals	4,592	60.7	4,807	60.6
VA hospitals	609	60.8	310	55.8
Private mental hospitals	**	**	1,255	74.7
General hospitals	2,849	65.4	938	78.3
Outpatients clinics	6,172	74.8	8,212	64.7

\*Counts of persons are unduplicated within each type of facility, but some duplication of individuals exists between types of facility.

\*\*Data on private mental hospital patients in Louisiana were not available for matching.

TABLE 4

Percent of Matching Census Schedules Found for  
Persons in the 10 Percent Sample\*, Louisiana and Maryland,  
by Relationship to Household Head

Relationship To Household Head	Louisiana		Maryland	
	Total in Sample	Percent Matched	Total in Sample	Percent Matched
Total	1,228	69.9	1,025	68.3
Head of Household	287	77.0	239	72.8
Wife of head	279	76.0	182	78.6
Child of head	382	74.9	296	72.6
Total immediate family of head	948	75.8	717	74.2
Other relatives of head	86	62.8	70	57.1
Non-relatives and persons living alone	161	39.8	165	49.1
Inmates	32	62.5	73	64.4

TABLE 5

Percent of Matching Census Schedules Found for \*  
Persons Aged 25 and Over in the 10 Percent Sample,  
Louisiana and Maryland, by Marital Status

Marital Status	Louisiana		Maryland	
	No. in Sample	Percent Matched	No. in Sample	Percent Matched
Total	851	68.5	695	67.1
Married	491	77.4	376	76.1
Widowed	59	57.6	59	61.0
Divorced	83	51.8	50	46.0
Separated	103	56.3	89	57.3
Never married	106	55.7	113	55.8

TABLE 6

Percent of Matching Census Schedules Found for  
Persons Aged 25 and Over in the 10 Percent Sample \*  
Louisiana and Maryland, by Educational Level

Education	Louisiana		Maryland	
	No. in Sample	Percent Matched	No. in Sample	Percent Matched
Total	851	68.5	694	67.1
None	35	68.6	14	71.4
Elementary	387	70.0	265	63.0
High School	294	65.0	279	72.4
College	97	70.1	87	64.4
Unknown	38	76.3	49	63.3

\* Random sample of admissions

TABLE 7

Percent of Matching Census Schedules Found for Persons  
Aged 25 and Over in the 10 Percent Sample,\*  
Louisiana and Maryland, by Sex and Employment Status

Sex and Employment Status	Louisiana		Maryland	
	No. in Sample	Percent Matched	No. in Sample	Percent Matched
<b>Males</b>	426	66.2	372	63.4
Working	185	69.7	159	69.8
Looking for work	92	53.3	64	48.4
Unable to work	92	68.5	77	63.6
Inmate	10	60.0	24	62.5
Other	34	85.0	25	64.0
Unknown	13	46.2	23	60.9
<b>Females</b>	425	70.8	322	71.4
Working	73	67.1	57	71.9
Looking for work	15	53.3	11	72.7
Keeping house	276	75.4	154	77.3
Unable to work	36	55.6	42	54.8
Inmate	8	37.5	26	57.7
Other	8	75.0	16	81.3
Unknown	9	77.8	16	68.8

\*Random sample of admissions

TABLE 8

Estimated Census Net Undercount, by Sex,  
Color and Age, United States, 1960\*\*

Age	Total	Percent Undercount			
		White Male	White Female	Non-white Male	Non-white Female
Total	2.3	1.1	1.7	10.3	7.1
Under 5	2.6	2.1	1.4	7.9	6.4
5-14	2.1	2.3	1.3	4.9	3.8
15-24	4.0	3.3	2.3	13.9	9.5
25-44	2.6	2.2	0.7	16.0	6.2
45-64	2.3	0.2	1.8	13.0	12.8
65 and over	0.9	(8.1)*	4.5	(7.9)*	2.6

\*Overcount

\*\*Source: Taeuber, C. and Hansen, M.H.: A preliminary evaluation of the 1960 Censuses of population and housing, Bureau of the Census. U.S. Department of Commerce, September, 1963, unpublished.



TABLE 9

A: Ratio  $\frac{m_1 e_2}{m_2 e_1}$  for White-Non-white and Male-Female Comparisons by State

	Louisiana				Maryland			
	Nw/W		M/F		Nw/W		M/F	
	Male	Female	White	Non-white	Male	Female	White	Non-white
Total	.985	.953	.927	.958	.877	.854	.919	.944
Under 18	1.027	.920	1.029	1.148	.953	.965	1.001	.988
18-24	1.109	.988	.930	1.044	.932	1.003	1.017	.947
25-44	1.084	.957	.871	.987	.899	.788	.868	.992
45-64	1.023	1.027	.885	.882	.879	.927	.876	.830
65+	.758	.963	.809	.637	.839	.912	.803	.667

B: Ratio of Enumeration Rates  $e_2/e_1$  for White-Non-white and Male-Female Comparisons, United States

	Nw/W		M/F	
	Male	Female	White	Non-white
Total	1.103	1.058	.994	1.036
Under 18	1.049	1.041	1.009	1.017
18-24	1.123	1.080	1.010	1.051
25-44	1.164	1.059	1.015	1.117
45-64	1.147	1.126	.984	1.002
65+	1.002	.980	.883	.903