USING THE MEDICARE CURRENT BENEFICIARY SURVEY FOR ANALYSIS OF RURAL HEALTH POLICY ISSUES

Working Paper No. 52

WORKING PAPER SERIES

North Carolina Rural Health Research Program

Cecil G. Sheps Center for Health Services Research
The University of North Carolina at Chapel Hill

725 Airport Road, CB #7590, Chapel Hill, N.C. 27599-7590
phone: 919/966-5541 fax: 919/966-5764

Sheps Center World Wide Web Address: www.shepscenter.unc.edu
NCRHRP address: www.unc.edu/research_programs/Rural_Program/rhp.html

USING THE MEDICARE CURRENT BENEFICIARY SURVEY FOR ANALYSIS OF RURAL HEALTH POLICY ISSUES

December 1997

Sally C. Stearns, Ph.D.
Rebecca T. Slifkin, Ph.D.
Thomas Walke, MSPH

North Carolina Rural Health Research Program

Cecil G. Sheps Center for Health Services Research
The University of North Carolina at Chapel Hill

Funding for this analysis was provided by the Federal Office of Rural Health Policy under Contract Number HRSA 93-857(P). Thomas Ricketts gave conceptual guidance, and Patricia Taylor, Claudia Shur and Marc Berk provided very useful comments. Ann Howard is responsible for substantial programming assistance. The authors are responsible for all opinions and any remaining errors.

OVERVIEW

The federal government conducts a number of national surveys for the purpose of assessing the current health care system. One example of such a survey is the Medicare Current Beneficiary Survey (MCBS), which is a longitudinal survey of a nationally representative sample of approximately 14,400 aged and disabled persons eligible for Medicare. The MCBS provides an assessment of the health care status, needs, access to services, and use of services by the Medicare population. In addition, the survey can be used to model the effects of possible policy changes.

Recently, a number of rural health researchers have used or expressed interest in using the MCBS to address issues of importance to rural health policy. While the survey as a whole is nationally representative, use of the MCBS for sub-group analysis may be problematic. Based on the sampling design, the only subgroups for which the survey is considered to be nationally representative are selected age groups. Therefore, national estimates for other subgroups such as rural populations in selected areas, are not guaranteed to be statistically valid. This is *not* to say that the survey can not be used to estimate, for example, the probability of hospitalization in non-metropolitan areas versus metropolitan areas, as it certainly can be used for such a purpose. However, for descriptive analyses of increasingly finer population categories, the estimates generated may not reflect the true population in those categories because of differences between that population and the sample that was selected for broader sampling goals (e.g., with oversampling of the disabled or very elderly). Furthermore, even though the survey can still be appropriately used in multivariate analyses to assess relationships between characteristics such as urbanicity of county of residence and outcomes of interest, the ability to obtain relatively precise estimates of such relationships depends on the distribution of the sample with respect to the characteristics.

This monograph provides an assessment of the use of the MCBS for addressing rural health policy issues. We consider the distribution of the sample across a number of geographic indicators commonly used by rural health researchers. In addition to the Office of Management and Budget's classification of counties as either metropolitan or non-metropolitan, we also consider the representativeness of the MCBS with respect to the following: census division and metropolitan distinction within census division, three measures of county urbanicity (percent urbanized, the Parker-Ghelfi Urban Influence Code, and the Economic Research Services (ERS) Rural-Urban Continuum Code), and two policy-relevant county designations (Health Provider Shortage Areas and Medically Underserved Areas). The monograph starts with a brief description of the MCBS and the distribution of the sample across metropolitan and non-metropolitan areas. The second section shows the distribution of the sample across commonly used geographic indicators of region, urbanicity, or health service availability. The third section demonstrates possible constraints on the use of the MCBS in making national estimates of health service use by these indicators. The paper concludes with a discussion of the strengths and weaknesses in using the MCBS for descriptive and multivariate analyses.

DESCRIPTION OF THE MCBS

Adler (1984) and Laschober and Olin (1996) provide detailed descriptions of the MCBS, including the sampling strategy, the survey instruments, the timing and repeat nature of the surveys, the samples used in different analysis files, and available public use files. Users of the MCBS should consult both of these excellent sources. The key points about the public use files and sampling strategy that are relevant for this monograph are summarized below.

Public Use Files

Two main public use files with a wide range of sociodemographic and health status measures are constructed for each year of the MCBS: (1) Cost and Use and (2) Access to Care. The samples and purposes of these two files are different. The Cost and Use file contains data for any individual who was surveyed at any point during the year, regardless of whether they had been followed for the entire year, and so includes the "ever-enrolled" population. The Cost and Use file

contains survey data for all medical services used in the calendar year (including expenditures and payment sources), with linkages to Medicare claims. Therefore, the Cost and Use file is the primary file for detailed analysis of health service use. The Access to Care file contains survey data (including some data on access to and satisfaction with care that are not available on the Cost and Use file) for the subset of persons in the Cost and Use file who were continuously enrolled (or "always enrolled") in Medicare and were living in households (i.e., in the community) during the calendar year represented by the file. Therefore, persons who died, became newly eligible for Medicare, or were institutionalized during the full year are not represented on the Access to Care file.

It is important to note that although we assess the MCBS sample distribution with respect to a number of measures of urbanicity and of county designations (e.g., HPSA and MUA), the only geographic indicators that are available on the Public Use files are the metropolitan status and census division. County Social Security Administration codes, which can be used to merge on county-level geographic measures, may be obtained for approved research purposes through the Health Care Financing Administration.

This monograph uses data from the 1992 Cost and Use file, which has a sample size of 13,039. (We conducted all of the analyses with the 1992 Access to Care file also, but the results are so similar that we present only the Cost and Use results.) While the specific numbers and estimates pertain only to 1992, our expectation is that the main points and general distribution of the sample are probably pertinent for subsequent years of the survey (at least until there is a major change in sample design).

Sampling Strategy

The target population for the survey consists of all persons enrolled in Medicare Part A or Part B (or both) who live in households or long-term care facilities in the United States and Puerto Rico. The sample was selected using a stratified, multistage area probability sample design involving: (1) selection of 107 primary sampling units (PSUs) which were either metropolitan statistical areas or clusters of nonmetropolitan counties; (2) selection of zip code clusters within PSUs; and (3) selection of beneficiaries within the sampled zip code cluster. The sample was selected to be nationally representative of the following age groups: under 45, 45-64, 65-69, 70-74, 75-79, 80-84, and 85 and over. In addition, disabled persons under age 65 and persons age 80 and over were oversampled to ensure that annual samples have enough persons with long-term care facility stays to produce reliable estimates.

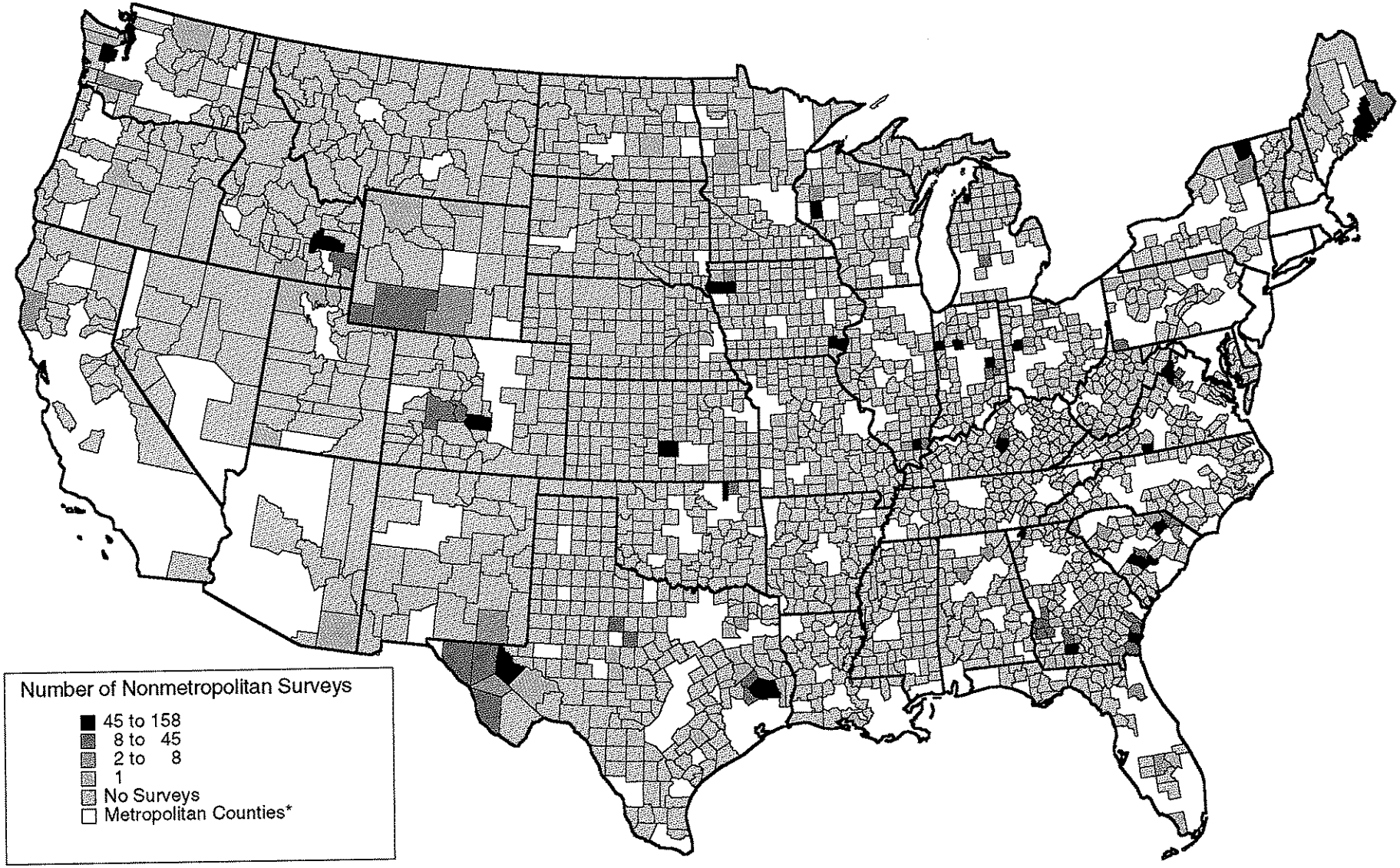
The sampling strategy has implications for the usefulness of the survey for rural policy analysis. Excluding persons who live in Puerto Rico and those for whom no county SSA code was available, the 1992 Cost and Use sample consists of 9,333 persons living in metropolitan areas and 3,486 persons living in non-metropolitan areas. However, the non-metropolitan sample comes from a limited number of geographic areas, as shown in Exhibit 1. Four states (Alaska, Delaware, Hawaii, and Montana) had no sample members selected. Among states with non-metropolitan counties, no persons were selected in non-metropolitan areas in four states (Massachusetts, Maryland, Nevada, and New Hampshire). Nineteen other states have fewer than 10 sample members in non-metropolitan counties. Although these exclusions are consistent with the goal of a nationally representative sample, persons interested in rural health policy must remember that certain areas (including some rural areas) are not directly represented by the survey sample. For example, policy-makers in Montana or Nevada can use data from the MCBS to predict health needs or the effect of policies on health service use in their state, but they may need to assume that the relationships that hold for nonmetropolitan areas in the rest of the United States also apply to them.

MCBS SAMPLE BY GEOGRAPHIC AREA INDICATORS

This section provides information on the distribution of the MCBS sample for the 1992 Cost and Use file by a number of geographic indica-

Exhibit 1

Sample Distribution: Medicare Current Beneficiary Survey, 1992



*Metropolitan counties are aggregated into white areas on the map.

Source: Medicare Current Benficiary Survey, Health Care Financing Administration, Office of the Actuary, Office of National Health Statistics, 1992.
Produced by: North Carolina Rural Health Research and Policy Analysis Center, Cecil G. Sheps Center for Health Services Research, University of North Carolina at Chapel Hill.

tors. Detailed descriptions of these indicators can be found in *What is 'Rural' and How to Measure 'Rurality,'* a technical issues paper prepared for the Federal Office of Rural Health Policy (Ricketts and Johnson-Web, 1997). In reviewing the distribution of the sample for each indicator, we consider both the number of counties from which the sample is drawn and the sample size (both in total and for persons age 65 or older). While the distribution of persons is important for characteristics which are measured at the individual level, the distribution across counties is also of interest, as researchers often assign county characteristics to the individual who resides in that county. We exclude persons in Puerto Rico from the analysis, since Puerto Rico is not relevant for most rural health policy issues in the U.S. To assess whether the sample distribution with respect to a given sub-group is significantly different from the overall distribution, we conduct two t-tests of the unweighted difference in proportions: the proportion of counties surveyed in each sub-group is compared to the proportion of total U.S. counties represented in the survey, and the proportion of the population 65 years and older sampled in each sub-group is compared to the proportion of the total U.S. population 65 years and older which is sampled.

Table 1 presents this information for the Office of Management and Budget's county metropolitan/non-metropolitan designation. The total MCBS sample of 12,819 individuals was drawn from 561 counties, representing 18.2 percent of all counties. As shown in the first two columns, non-metropolitan counties are significantly *underrepresented* (meaning that there are proportionately fewer of them in the sample compared to the overall average), while metropolitan counties are significantly *overrepresented* by the sampling scheme. These differences occur as a result of the survey design, which is intended to be representative of the population rather than counties. Non-metropolitan counties are less well represented because they are less densely populated.

In contrast to the county representation, the percent of population age 65 and over represented is actually greater at 0.0369 percent for non-metropolitan areas and lower at 0.0296 percent for metropolitan areas than for the sample as a whole (at 0.0313 percent), though the differences are not statistically significant for either category. Therefore, the sample within the metropolitan and non-metropolitan categories may be reasonably representative of the population within those categories: roughly 27 and 73 percent of sample individuals lived in a non-metropolitan and metropolitan counties, respectively, for both the total sample and the aged sample of 10,482. Although the percentage of sampled individuals across metropolitan and non- metropolitan counties is not significantly different than for the U.S. as a whole, this assessment does not address the representativeness of the sample for characteristics of the rural population such as age and sex, nor does it address the representativeness of the primary sampling units selected. Furthermore, as will be developed shortly, the absolute sample size may provide the biggest limitation for sub-group analysis by selected geographic indicators.

Many of the above points hold for other geographical indicators. With respect to census division (see Table 2), the analysis shows that the percent of counties in the sample varies considerably across the different regions. The percent of population represented by the sample does not vary significantly across the census divisions from the total, though the number of observations within a division ranges from 409 to 2,681 (or from 340 to 2,141 for the aged sample). Further breakouts by census division and metropolitan indicator (Table 3) reinforce these patterns.

Three geographic indicators commonly used by rural health researchers to define rural areas are the percent of a county that is urbanized, Urban Influence Codes, and the ERS Rural-Urban Continuum Code. Table 4 contains the distribution of the MCBS sample with respect to these county-level indicators. As with the other geographic indicators, the percent of counties in each sub-category which are sampled often varies significantly from the percent of counties found in those categories across the entire United States. With respect to the percent urbanized indicator, counties with less than 40 percent of the population urbanized were selected at a lower rate than the overall sample, while counties with

Table 1: 1992 MCBS Cost and Use Sample by Metropolitan Area

	MCBS Number of Counties	Percent of of Counties in the Sample	MCBS Number of Individuals	MCBS Number of Individuals 65 and Over	Percent of Population 65 and Over in the Sample
Nonmetropolitan	202	9.0% (* -)	3,486	2,827	0.0369%
Metropolitan	359	43.1% (* +)	9,333	7,655	0.0296%
Total or Overall Average	561	18.2%	12,819	10,482	0.0313%

Table 2: 1992 MCBS Cost and Use Sample by Census Division

	MCBS Number of Counties	Percent of of Counties in the Sample	MCBS Number of Individuals	MCBS Number of Individuals 65 and Over	Percent of Population 65 and Over in the Sample
New England	18	26.9% (* +)	409	340	0.0183%
Middle Atlantic	75	50.0% (* +)	2,290	1,887	0.0348%
East North Central	94	21.5%	2,169	1,775	0.0319%
West North Central	50	8.1% (* -)	837	692	0.0274%
South Atlantic	133	23.9% (* +)	2,681	2,141	0.0332%
East South Central	41	11.3% (* -)	762	576	0.0284%
West South Central	73	15.5%	1,321	1,105	0.0344%
Mountain	33	11.7% (* -)	737	611	0.0346%
Pacific	44	31.9% (* +)	1,613	1,355	0.0289%
Total or Overall Average	561	18.2%	12,819	10,482	0.0313%

Table 3: 1992 MCBS Cost and Use Sample by Metropolitan Indicator and Census Division

	MCBS Number of Counties	Percent of of Counties in the Sample	MCBS Number of Individuals	MCBS Number of Individuals 65 and Over	Percent of Population 65 and Over in the Sample
Nonmetropolitan					
New England	4	13.3% (* -)	135	117	0.0648%
Middle Atlantic	10	17.2%	113	88	0.0175%
East North Central	36	12.5% (* -)	615	519	0.0391%
West North Central	25	4.6% (* -)	413	345	0.0274%
South Atlantic	55	15.4%	1,133	874	0.0601%
East South Central	20	7.0% (* -)	168	115	0.0124%
West South Central	27	7.5% (* -)	496	424	0.0413%
Mountain	17	7.0% (* -)	250	205	0.0395%
Pacific	8	9.9% (* -)	163	140	0.0300%
Total Nonmetropolitan	202	9.0% (* -)	3,486	2,827	0.0369%
Metropolitan					
New England	14	37.8% (* +)	274	223	0.0133%
Middle Atlantic	65	70.7% (* +)	2,177	1,799	0.0365%
East North Central	58	38.9% (* +)	1,554	1,256	0.0296%
West North Central	25	34.2% (* +)	424	347	0.0274%
South Atlantic	78	39.0% (* +)	1,548	1,267	0.0254%
East South Central	21	26.6% (* +)	594	461	0.0419%
West South Central	46	42.6% (* +)	825	681	0.0312%
Mountain	16	42.1% (* +)	487	406	0.0325%
Pacific	36	63.2% (* +)	1,450	1,215	0.0287%
Total Metropolitan	359	43.1% (* +)	9,333	7,655	0.0296%

* representation for the sample relative to the overall average for the nation is significantly greater (+) or less (-) with 95 percent confidence

at least 60 percent of the population urbanized were relatively more likely to be selected. Similarly, when assessing the sample distribution relative to Urban Influence Codes (Table 4b), metropolitan counties are significantly overrepresented while all nonmetropolitan counties (with the exception of those adjacent to small metropolitan counties and having a city with population greater than 10,000) are significantly underrepresented. The percent of the population represented by the sample reflects two particular problems that are also revealed by the absolute sample size. Specifically, two of the urban influence codes (for counties adjacent to a large metropolitan area and having a city greater than 10,000, and for counties not adjacent to a metropolitan area and having a city less than 2,500) have very few observations. From an analytic perspective, most analyses would have to collapse these codes into other categories because of the extremely small sample. Some analysts might find it useful to collapse additional categories. For example, Dansky and colleagues (1997) tried to maximize the statistical power of their analysis by collapsing the ten category Human Resource Profile County Adjacency Code, which is based on county population size and adjacency to metropolitan areas, into five categories.

Similar points may be made about the ERS Rural Urban Continuum Code (Table 4c). The problems of small sample size and proportion of the population represented are not quite as extreme for counties that are fringe counties for metropolitan areas. However, the problem persists for non-metropolitan counties that are not adjacent to metropolitan areas and have urban population less than 2,500. Also, the sample sizes are under 300 for two other categories of counties: non-adjacent non-metropolitan counties with urban population greater than 20,000, and non-metropolitan counties that are adjacent to a metropolitan county but have less than 2,500 urban population (though the proportion of population represented is higher than for the sample on average for the latter category).

The same patterns of percent of counties in the sample and percent of population in the sample are also found for the two policy-relevant county designations: Health Provider Shortage Areas (1995 designation), and Medically Underserved Areas (1981 designation). The distribution of the percent of the population in the sample in each sub-category does not differ significantly from the percent of total U.S. population in the survey sample. However, the fact that only 88 individuals were sampled from whole county metropolitan HPSAs limits the ability to do any descriptive analysis with this category or comparisons with whole county nonmetropolitan HPSAs. In contrast, the sample sizes are at least moderate for the three HPSA or MUA designations (none, whole county or partial county) for non-metropolitan areas, though the percent of counties is under-representative relative to the study as a whole.

NATIONAL ESTIMATES OF HOSPITAL USE BY GEOGRAPHIC AREA INDICATORS

While the prior section illustrated some of the problems inherent in the distribution of the MCBS sample with respect to particular geographic categorizations (such as non-adjacent non-metropolitan counties with urban population less than 2,500), this section takes an alternative analytic approach to demonstrating some of the issues that occur in using the MCBS for analysis of rural health policy issues. Given concerns about access to care in rural areas, the MCBS may be used, for example, to analyze whether the probability of hospitalization differs between different areas. We therefore proceed to use the 1992 MCBS Cost and Use file to estimate the probability of an individual having one or more hospital stays in 1992, stratified by the various geographic indicators employed in the prior section. In addition to providing an estimate of this probability for each of the geographic areas, we also provide a 95 percent confidence interval for each estimate. Since the MCBS is based on a complex survey design (which involves correlations between the observations due to the sampling strategy), we use bootstrap techniques to obtain the confidence intervals (Efron and Tibshirani, 1993). The bootstrap approach draws

Table 4: 1992 MCBS Cost and Use Sample by Urban Indicators

Table 4a: Percent Urbanized

	MCBS Number of Counties	Percent of of Counties in the Sample	MCBS Number of Individuals	MCBS Number of Individuals 65 and Over	Percent of Population 65 and Over in the Sample
0-20	61	5.9% (* -)	724	609	0.0276%
21-40	105	14.5% (* -)	1,731	1,393	0.0400%
41-60	111	17.5%	1,505	1,191	0.0291%
61-80	96	24.4% (* +)	1,958	1,605	0.0322%
81-100	188	62.0% (* +)	6,901	5,684	0.0303%
Total or Overall Average	561	18.2%	12,819	10,482	0.0313%

Table 4b: Urban Influence Code

	MCBS Number of Counties	Percent of of Counties in the Sample	MCBS Number of Individuals	MCBS Number of Individuals 65 and Over	Percent of Population 65 and Over in the Sample
Metropolitan					
Large Metro (> 1 million)	189	62.4% (* +)	5,438	4,499	0.0291%
Small Metro (< 1 million)	165	32.2% (* +)	3,889	3,151	0.0308%
Nonmetropolitan					
Adj.to Large Metro; city > 10,000	9	14.5% (* -)	12	5	0.0010% (* -)
Adj. to Large Metro; city < 10,000	14	11.5% (* -)	339	284	0.0718%
Adj. to Small Metro; city > 10,000	36	19.6%	1,049	843	0.0583%
Adj. to Small Metro; city < 10,000	57	9.2% (* -)	919	772	0.0391%
Not adj. to Metro; city > 10,000	33	14.3% (* -)	474	371	0.0272%
Not adj. to Metro; 2,500 <> 9,999	49	9.0% (* -)	665	534	0.0348%
Not adj. to Metro; city < 2,500	9	1.8% (* -)	34	23	0.0039% (* -)
Total or Overall Average	561	18.2%	12,819	10,482	0.0313%

Table 4c: Economic Research Service Rural-Urban Continuum Code

	MCBS Number of Counties	Percent of of Counties in the Sample	MCBS Number of Individuals	MCBS Number of Individuals 65 and Over	Percent of Population 65 and Over in the Sample
Metropolitan					
Central > 1 million	138	82.1% (* +)	5,023	4,160	0.0294%
Fringe > 1 million	51	38.6% (* +)	415	339	0.0298%
250,000-1million	117	36.8% (* +)	2,937	2,361	0.0309%
< 250,000	48	24.2% (* +)	952	790	0.0288%
Nonmetropolitan					
Adjacent; > 20,000 urban pop	28	21.1%	754	586	0.0414%
Not adjacent; > 20,000 urban pop	18	16.4%	268	218	0.0257%
Adjacent; 2,500-19,999 urban pop	69	11.3% (* -)	1,336	1,124	0.0445%
Not adjacent; 2,500-19,999 urban	64	10.0% (* -)	871	687	0.0342%
Adjacent; < 2, 500 urban pop	19	7.7% (* -)	229	194	0.0484%
Not adjacent; < 2, 500 urban pop	9	1.7% (* -)	34	23	0.0037% (* -)
Total or Overall Average	561	18.2%	12,819	10,482	0.0313%

* representation for the sample relative to the overall average for the nation is significantly greater (+) or less (-) with 95 percent confidence

Table 5: 1992 MCBS by Health Provider Shortage Area (HPSA) in 1995

Table 5a: Overall

	MCBS Number of Counties	Percent of of Counties in the Sample	MCBS Number of Individuals	MCBS Number of Individuals 65 and Over	Population 65 and Over in the Sample
None	202	18.2%	3,055	2,536	0.0283%
Whole County	63	8.0% (*-)	878	694	0.0371%
Partial County	296	25.0% (* +)	8,886	7,252	0.0319%
Total or Overall Average	561	18.2%	12,819	10,482	0.0313%

Table 5b: By Metropolitan/Non-Metropolitan Indicator

	MCBS Number of Counties	Percent of of Counties in the Sample	MCBS Number of Individuals	MCBS Number of Individuals 65 and Over	Percent of Population 65 and Over in the Sample
Nonmetropolitan					
None	77	9.9% (* -)	1,225	998	0.0342%
Whole County	50	7.0% (* -)	790	626	0.0411%
Partial County	75	9.9% (* -)	1,471	1,203	0.0373%
Total Nonmetropolitan	202	9.0% (* -)	3,486	2,827	0.0369%
Metropolitan					
None	125	37.5% (* +)	1,830	1,538	0.0254%
Whole County	13	18.8%	88	68	0.0196%
Partial County	221	51.3% (* +)	7,415	6,049	0.0311%
Total Metropolitan	359	43.1% (* +)	9,333	7,655	0.0296%

Table 6: 1992 MCBS Cost and Use Sample by Medically Underserved Area Indicator (1981)

Table 6a: Overall

	MCBS Number of Counties	Percent of of Counties in the Sample	MCBS Number of Individuals	MCBS Number of Individuals 65 and Over	Percent of Population 65 and Over in the Sample
None	136	21.3%	2,507	2,095	0.0381%
Whole County	140	9.7% (*-)	2,011	1,619	0.0324%
Partial County	285	28.7% (*+)	8,301	6,768	0.0294%
Total or Overall Average	561	18.2%	12,819	10,482	0.0313%

Table 6b: By Metropolitan/Non-Metropolitan Indicator

	MCBS Number of Counties	Percent of of Counties in the Sample	MCBS Number of Individuals	MCBS Number of Individuals 65 and Over	Percent of Population 65 and Over in the Sample
Nonmetropolitan					
None	48	10.9% (*-)	1,177	962	0.0548%
Whole County	100	7.8% (*-)	1,435	1,140	0.0327%
Partial County	54	10.1% (*-)	874	725	0.0300%
Total Nonmetropolitan	202	9.0% (*-)	3,486	2,827	0.0369%
Metropolitan					
None	88	44.2% (* +)	1,330	1,133	0.0303%
Whole County	40	23.0% (* +)	576	479	0.0317%
Partial County	231	50.2% (* +)	7,427	6,043	0.0293%
Total Metropolitan	359	43.1% (* +)	9,333	7,655	0.0296%

* representation for the sample relative to the overall average for the nation is significantly greater (+) or less (-) with 95 percent confidence

repeated samples with replacement from the original sample to develop an empirical distribution of the estimate of interest. We present bias-corrected percentile intervals which account for any non-normality in the distribution of the variable of interest.[1]

Based on the MCBS, the estimated probability of hospitalization in 1992 in non-metropolitan counties is 0.184 (or 18.4 percent), which is greater than the probability of 0.175 in metropolitan counties (Exhibit 2). As demonstrated by the confidence intervals, however, it is not possible to say with 95 percent confidence that these estimates are statistically different from each other. There is a substantial amount of overlap between the two confidence intervals; each of the estimates is within the 95 percent confidence interval for the other category. The confidence interval for non-metropolitan areas (ranging from 0.171 to 0.199) is wider than the interval for metropolitan areas (0.167 to 0.183) largely because of the smaller number of observations (3,486 as compared to 9,333).

Exhibit 3 provides the estimates of the probability of hospitalization by census division, with the non-metropolitan and metropolitan breakouts in Exhibits 4 and 5. In both Exhibit 3 and Exhibit 4, the confidence intervals overlap for all the regions, precluding a definitive determination of differences in the overall probability of hospitalization by region (in total or for non-metropolitan counties within regions). The confidence intervals in Exhibit 5 are narrower than in Exhibit 4, once again primarily due to the larger sample sizes for the metropolitan as compared to non-metropolitan counties. Exhibit 5 does show that the estimated probability of hospitalization of 0.201 for individuals residing in metropolitan counties in the Middle Atlantic region (with confidence interval ranging form 0.184 to 0.218) is significantly higher than for the West North Central region probability of 0.133 (with confidence interval ranging from 0.100 to 0.169).

The confidence intervals overlap either substantially or entirely for the three measures of urbanicity (Exhibits 6-8). Some of the confidence intervals are so wide that the estimates are not very useful, clearly indicating the need to collapse categories for analytic purposes. In Exhibits 7 and 8, the categories with fewer than 40 observations have extremely wide confidence intervals and should be combined with other categories. Even with such a modification, however, the overlap between the categories precludes determination of any significant difference in the probability of hospitalization related to degree of urbanicity.

The same points can be made about the estimates according to the two policy-relevant variables (HPSA in Exhibit 9 and MUA in Exhibit 10). For example, in Exhibit 9, the estimated probability of hospitalization of 0.143 for persons in metropolitan whole county HPSAs falls below the lower 95 percent confidence interval for all the other area designations. However, the fact that the estimate is based on only 88 observations makes it impossible to feel confident that the probability of hospitalization truly differs for this designation; in fact, the upper confidence interval of 0.217 for this group is actually *higher* than the upper confidence interval for most of the other groups.

SUMMARY: USING THE MCBS FOR DESCRIPTIVE AND MULTIVARIATE ANALYSES

This paper provides information on the distribution of the 1992 Cost and Use file by several geographic indicators, including some that are not available with the MCBS Public Use files. Knowledge of these distributions may be helpful to rural health policy researchers contemplating use of the MCBS. We demonstrate that the non-metropolitan counties selected for the sample are a small proportion of total non-metropolitan counties (small at least relative to the overall sample), but that the proportion of the population in the sample is similar to that for the overall sample except for a few geographic designations.

We also demonstrate the impact of small

[1] Percentile confidence intervals are non-parametric intervals that have two advantages over the standard normal intervals. First, they allow for a non-asymmetric confidence interval, if appropriate. Second, they limit confidence intervals for the estimate of a probability to the 0,1 range.

sample sizes on the width of confidence intervals for survey estimates by conducting a bivariate descriptive analysis of the probability of hospitalization across the selected geographic indicators as an illustration. We show that small sample sizes within some of the designations may limit the ability to determine policy-relevant differences across the types of geographic areas, as one can not distinguish between true lack of substantive difference and imprecision due to sample size.

It is true that the probability of hospitalization is not the only policy-relevant measure of utilization for comparison across geographic designations. It is certainly conceivable that other measures such as length of stay or frequency of ambulatory care use might have patterns with greater statistically significant variation across the geographic categorizations used in this paper. Yet the probability of hospitalization would seem to be an important variable for policy purposes, so the knowledge that confidence intervals on at least this variable are too wide to allow for definitive determination of differences related to geographic location illustrates limitations in using the MCBS for descriptive analyses.

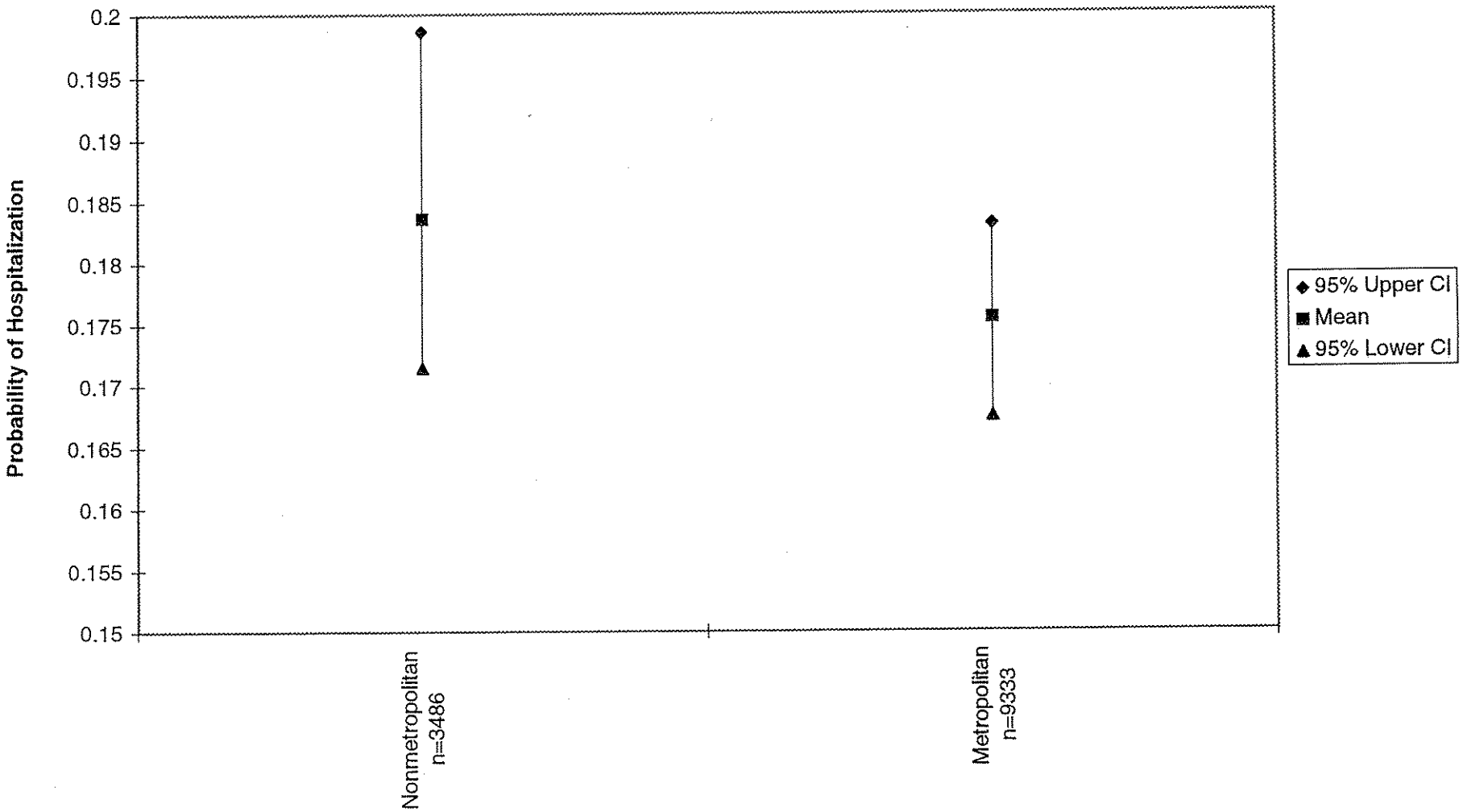
The most important point from this analysis pertains to the relevance of the MCBS for multiple regression analysis in which a number of variables are controlled for rather than descriptive bivariate analysis. Specifically, the analyses in this paper did not control for any other characteristics of the individuals. One advantage and two cautions regarding regression analysis should be noted.

Regarding the advantage of multiple regression analysis, simultaneously controlling for other factors (e.g., age, health status, income, etc.) is most likely critical for determining true differences in use measures such as the probability of hospitalization across geographic regions. For example, the fact that the probability of hospitalization was higher in non-metropolitan areas than in metropolitan areas (albeit not significantly so) may actually be attributed to underlying differences in the population. For example, non-metropolitan Medicare enrollees may be older, sicker, and poorer than metropolitan enrollees. It is conceivable that the differences in these characteristics are driving the observed difference, and it is only by controlling for these factors through multiple regression analysis or logistic regression that a more pure assessment of differences in the probability of hospital use attributable to the type of geographic area (or service availability within those areas) could be determined.

Multiple regression analysis, however, does not solve the problem of small sample size for a particular type of geographic area of interest. Confidence intervals for coefficients of geographic categories with relatively few observations in the data set will continue to be fairly large and imprecise, resulting in an inability to detect meaningful differences across categories. Therefore, collapsing categories for certain geographic indicators may still be appropriate. A more subtle (and less easily dealt with) concern pertains to the ability to use regression results to predict an outcome variable in relation to county characteristics if those counties are not represented in the sample. The analysis in this paper demonstrates that nonmetropolitan counties are underrepresented as a percent of total counties represented in the MCBS. This result may mean that the full spectrum of rural county characteristics *may* not be represented in the MCBS. Individual researchers should carefully review the distribution of county characteristics they wish to include in their models to ensure that the MCBS sample represents the full range of those characteristics observed nationally. Predictions for any characteristic that is not in the range represented by the MCBS sample would constitute "out-of-sample" predictions and would not be very reliable.

Although some rural policy issues may be of sufficient importance to argue for a different sampling design that provides proportional representation of counties, the decision to proceed with such a goal would have to be justified based on design and cost implications. Despite the cautions above, the MCBS is most likely a useful tool for many types of rural health policy analysis. The sample distribution by geographic characteristics means, however, that regional differences in access and service use can not be determined by bivariate assessments. Instead, fuller models of access and service use are needed.

Exhibit 2: Probability of Hospitalization by Metropolitan Area



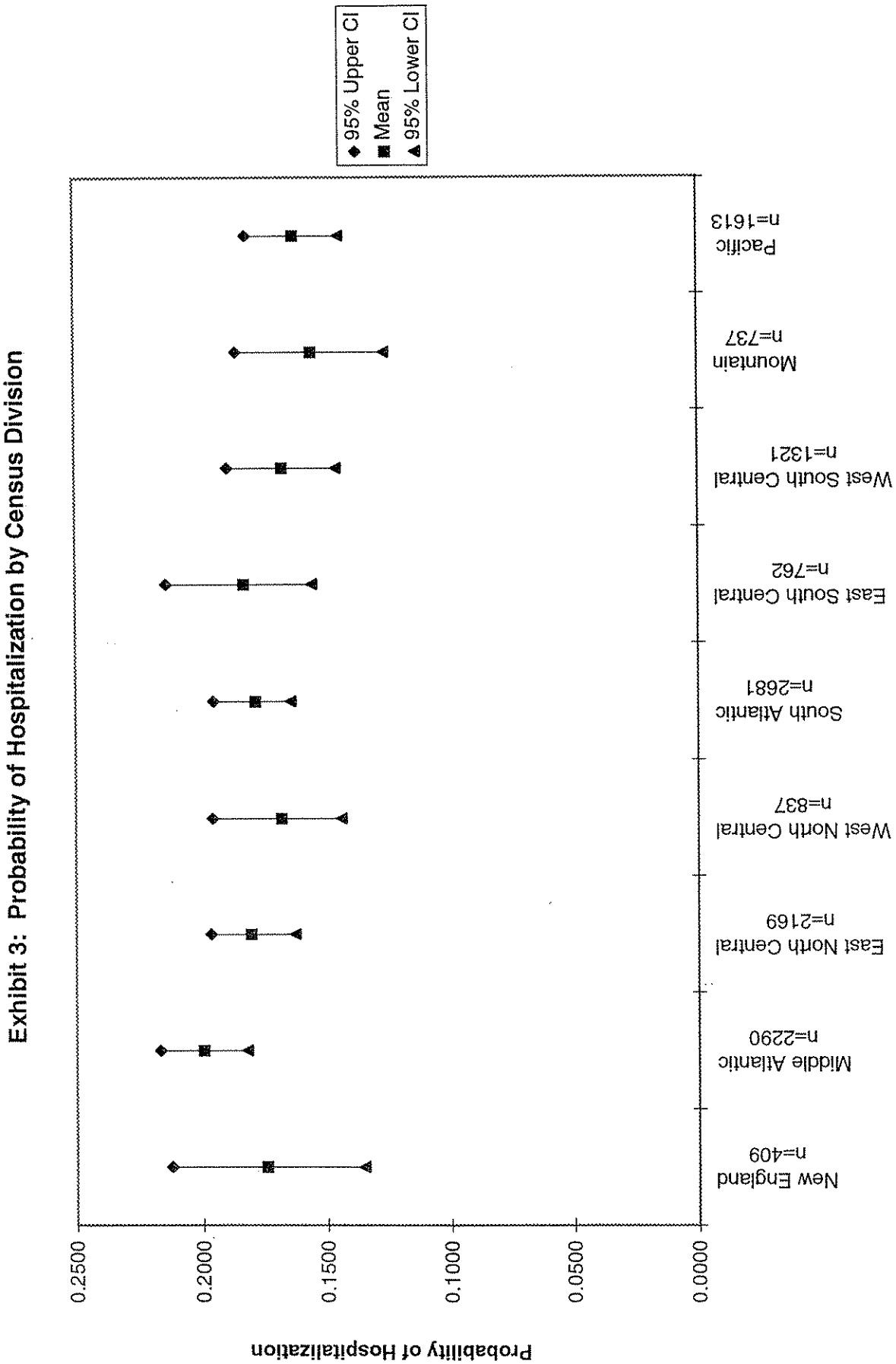


Exhibit 3: Probability of Hospitalization by Census Division
Probability of Hospitalization
0.2500
0.2000
0.1500
0.1000
0.0500
0.0000
95% Upper CI
Mean
95% Lower CI
New England n=409
Middle Atlantic n=2290
East North Central n=2169
West North Central n=837
South Atlantic n=2681
East South Central n=762
West South Central n=1321
Mountain n=737
Pacific n=1613

Exhibit 4: Probability of Hospitalization by Census Division & Nonmetro

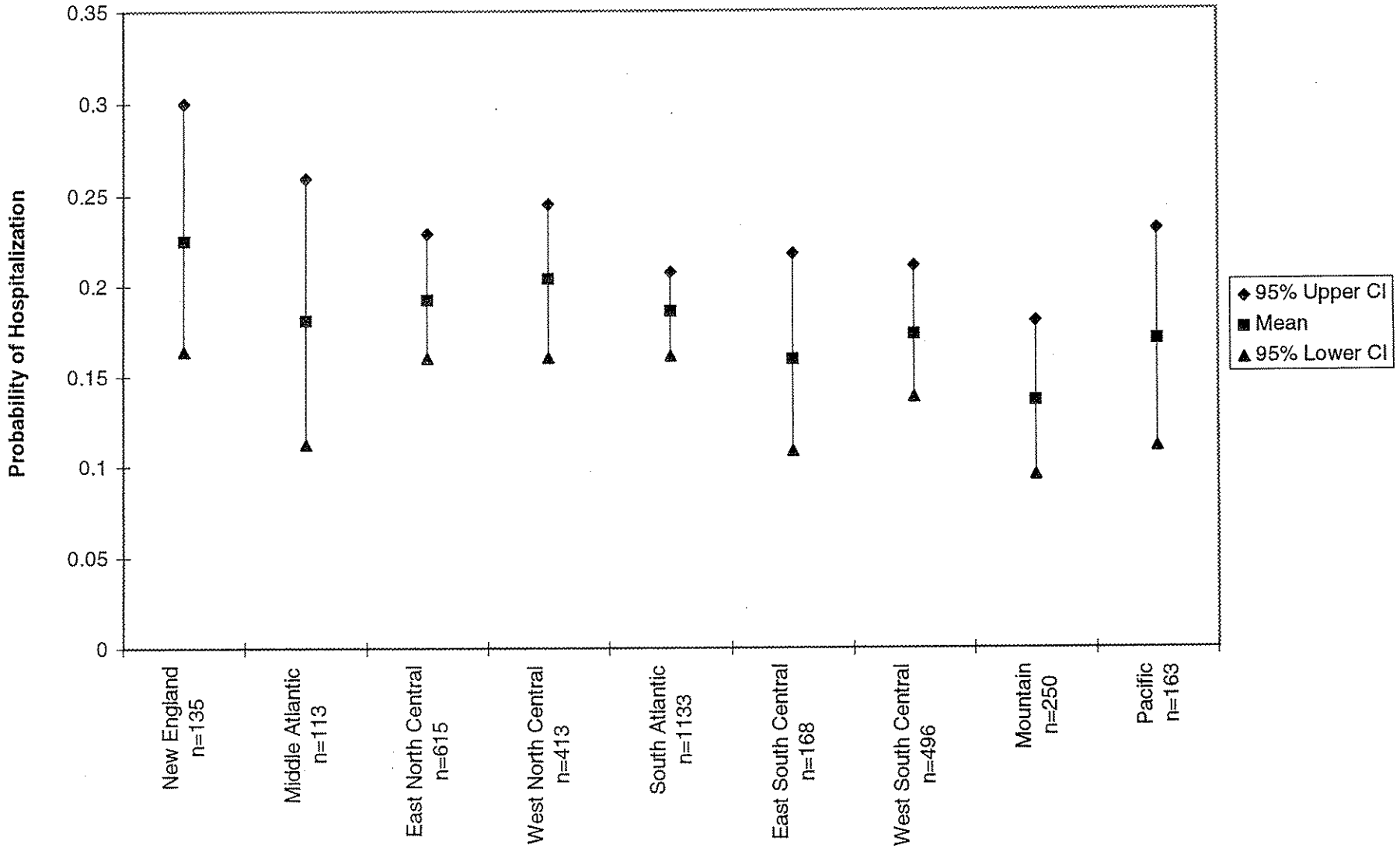


Exhibit 5: Probability of Hospitalization by Census Division & Metro

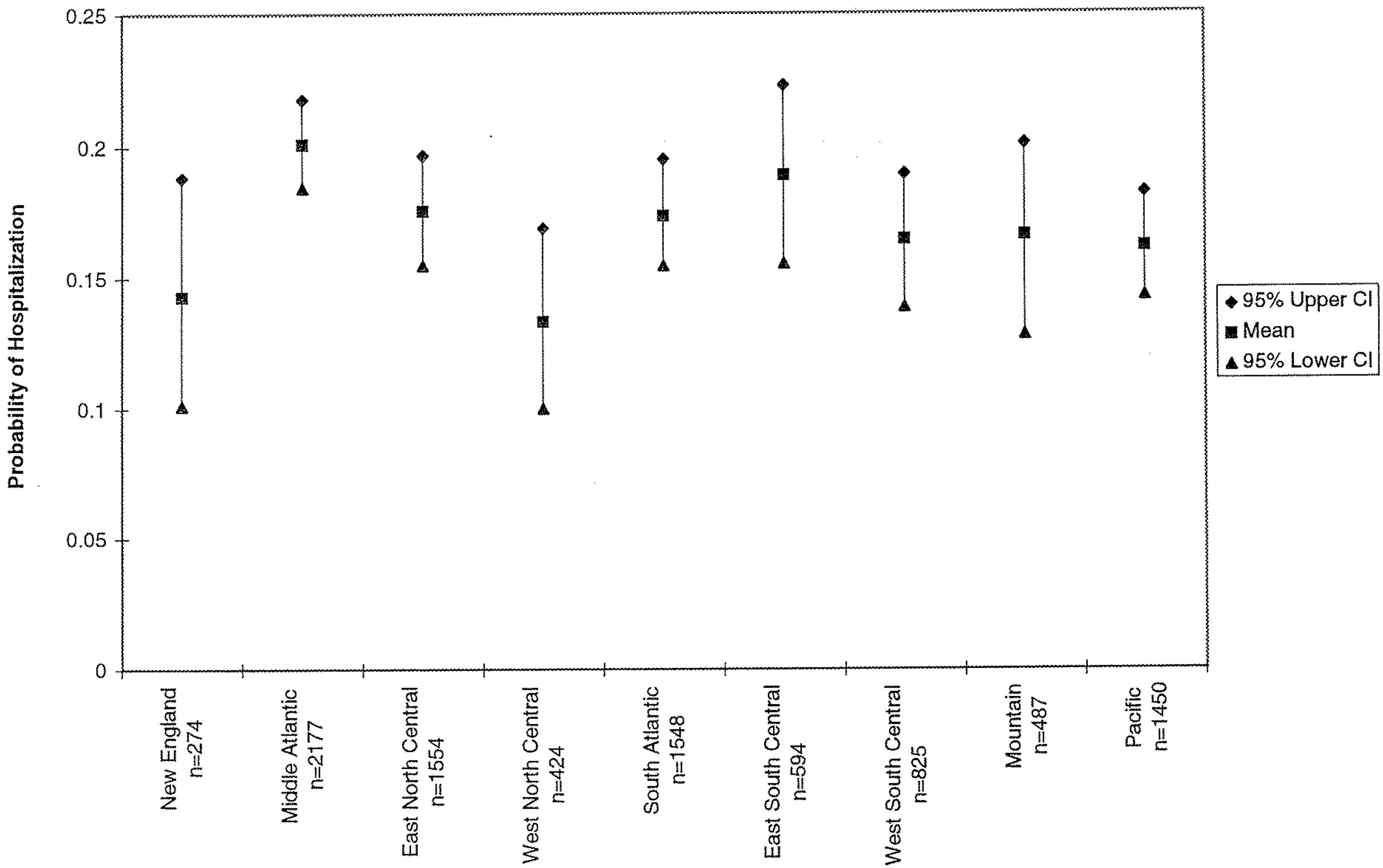


Exhibit 6: Probability of Hospitalization by Percent Urban

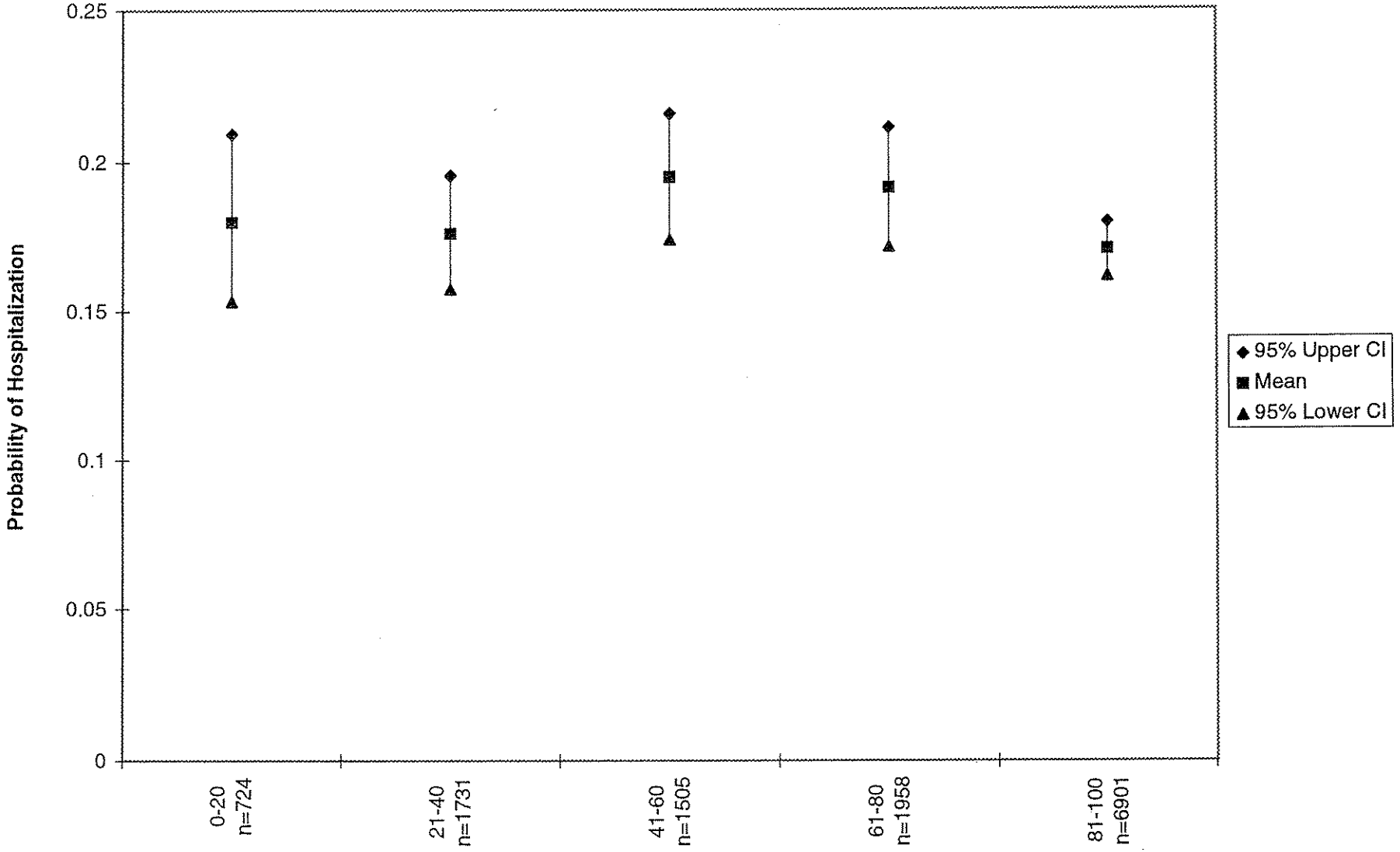


Exhibit 7: Probability of Hospitalization by Urban Influence Code

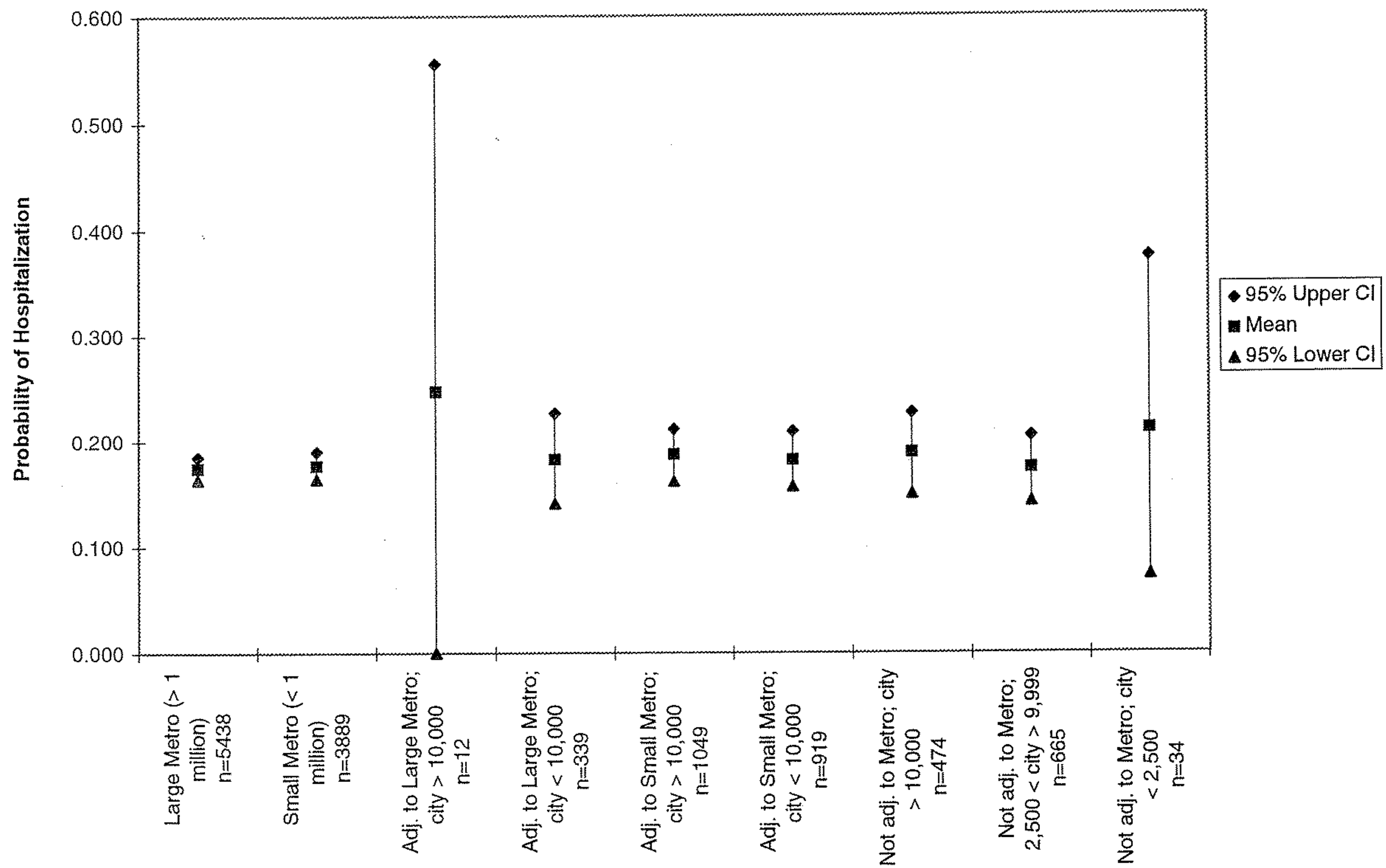


Exhibit 8: Probability of Hospitalization by ERS Rural-Urban Continuum Code

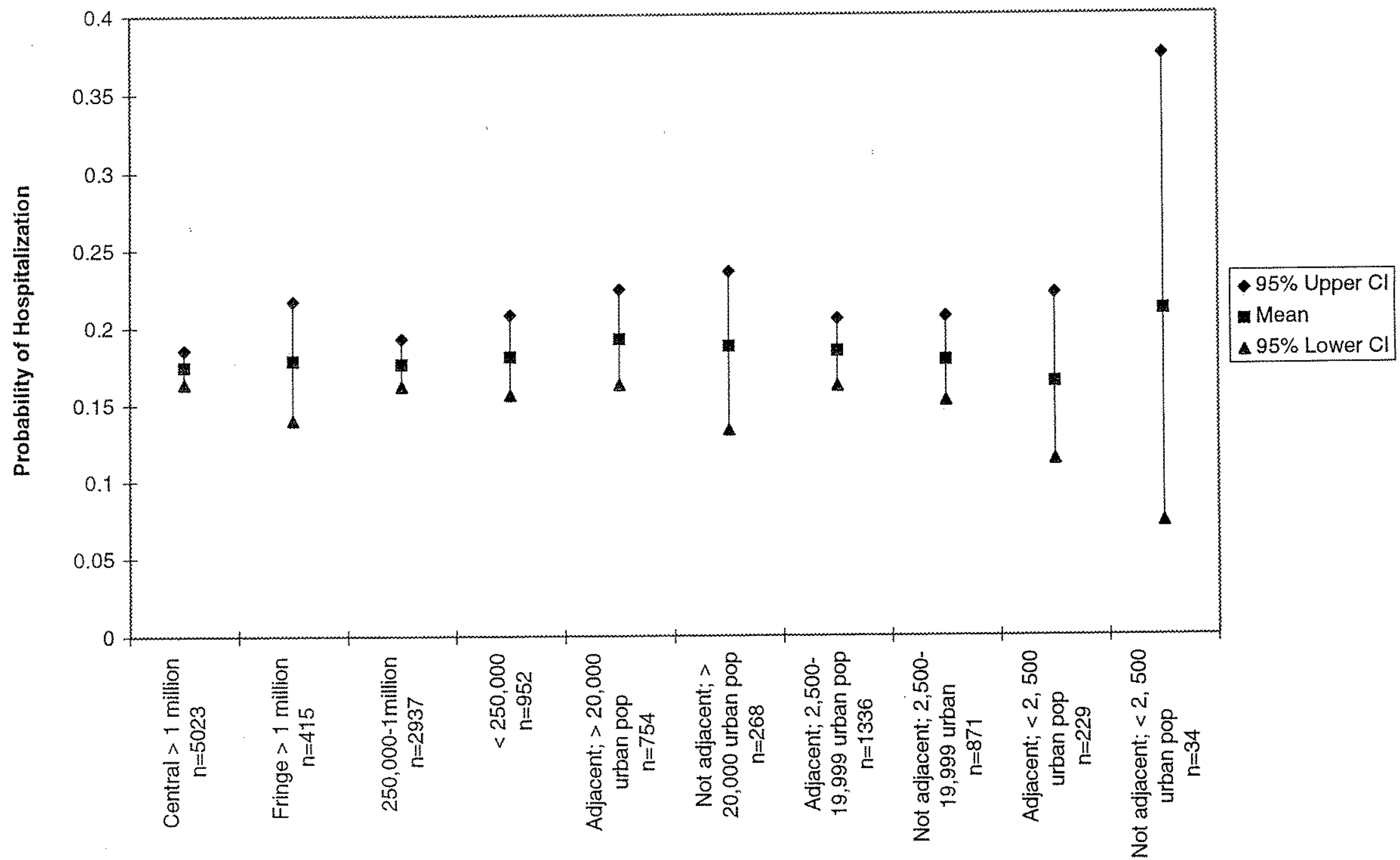


Exhibit 9: Probability of Hospitalization by HPSA 1995

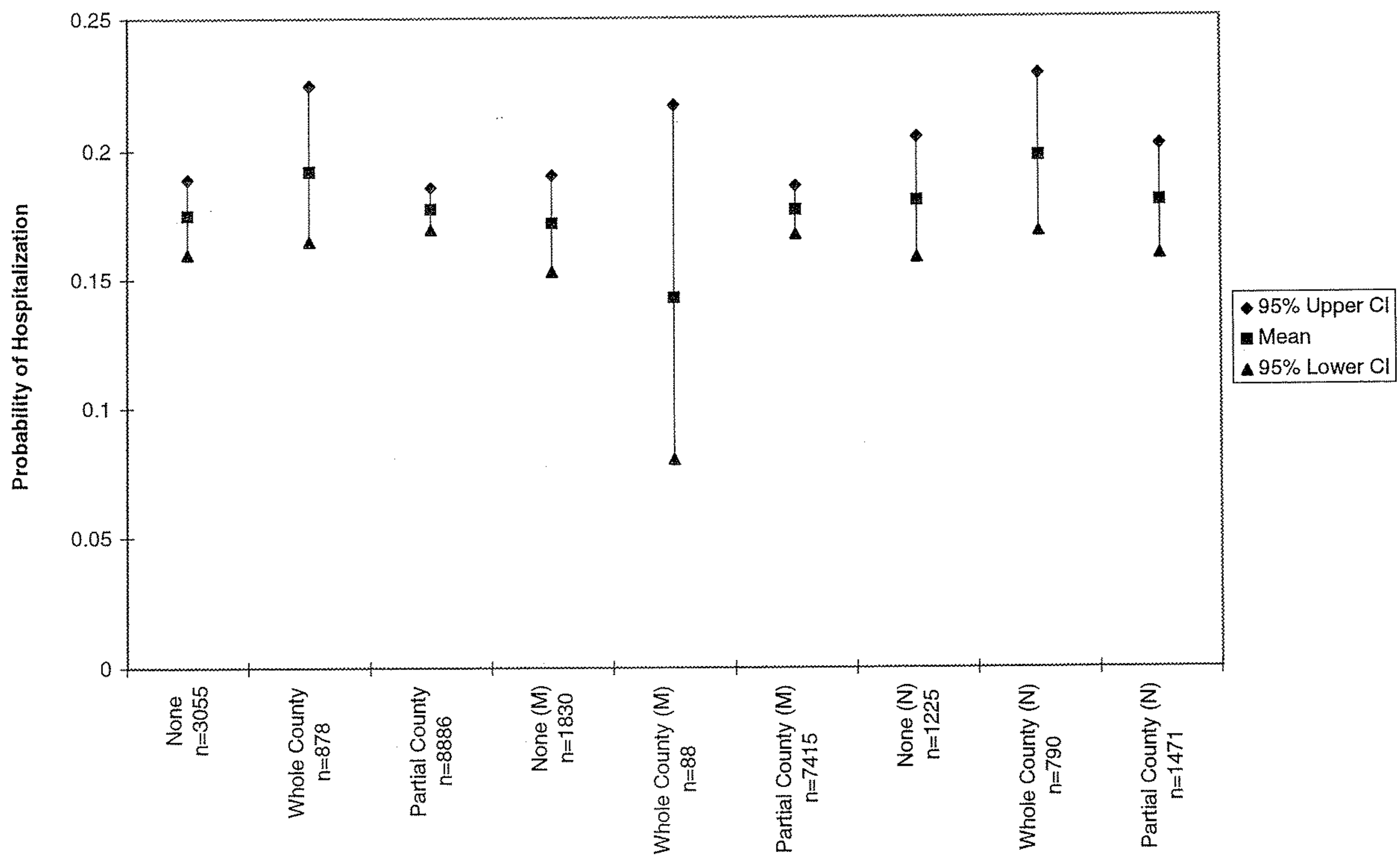
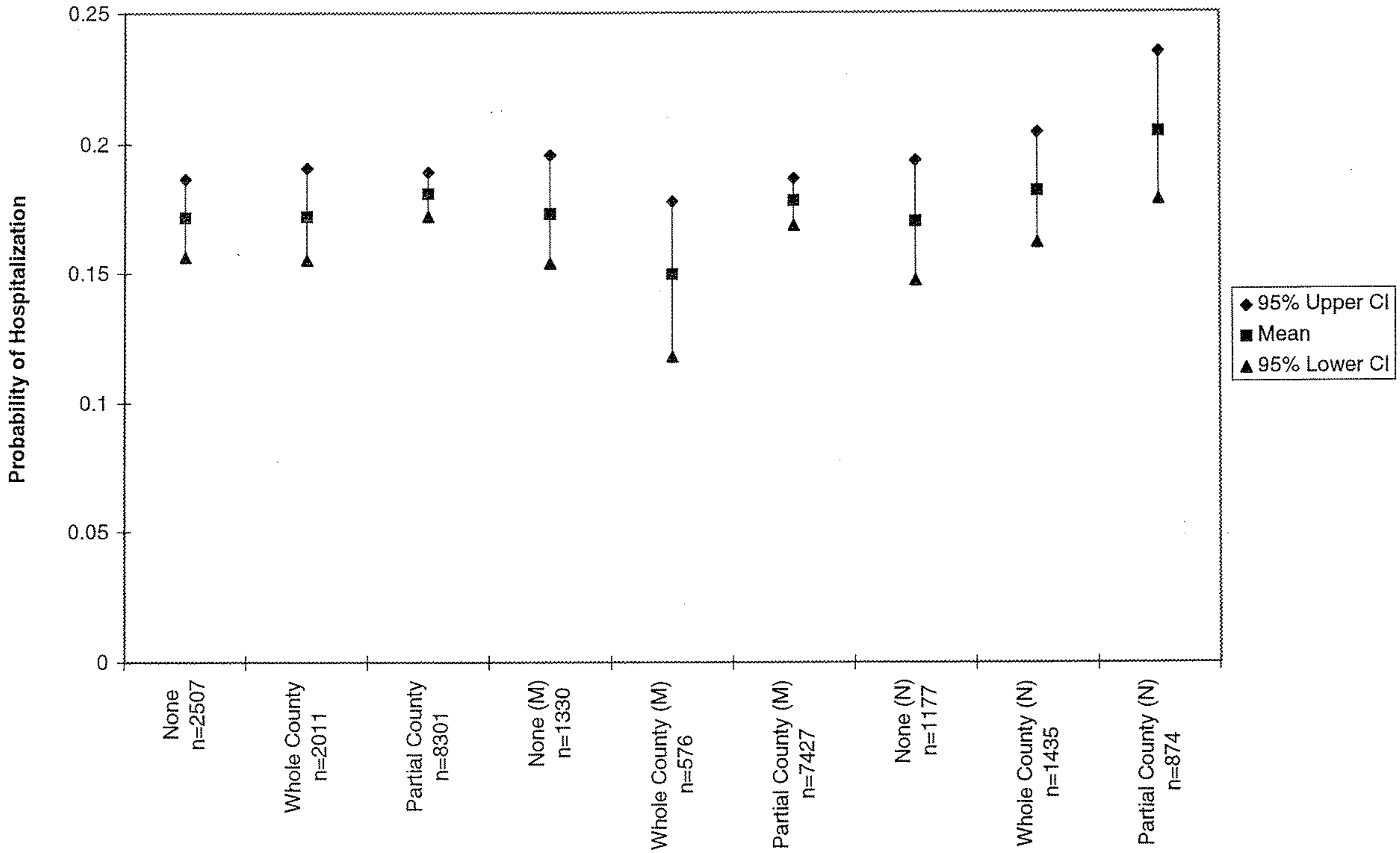


Exhibit 10: Probability of Hospitalization by MUA 1981



REFERENCES

Adler, G.S. (1994) "A Profile of the Medicare Current Beneficiary Survey." *Health Care Financing Review*, 15(4):153-163.

Dansky, K.H, D. Brannon, D.G. Shea, J. Vasey, R. Dirani. (1997) "Profiles of Hospital, Physician, and Home Health Service Use by Older Persons in Rural Areas." Paper under review at *The Gerontologist*.

Efron, B., and R.J. Tibshirani. (1993) *An Introduction to the Bootstrap*, Chapman and Hall, N.Y.

Laschober, M.A. and G.L. Olin (1996) *Health and Health Care of the Medicare Population: Data from the 1992 Medicare Current Beneficiary Survey*. Rockville, MD: Westat, Inc.

Ricketts, T., and K. Johnson-Webb. (1997) *What is 'Rural' and How to Measure 'Rurality': A Focus on Health Care Delivery and Health Policy*. Technical Issues Paper, North Carolina Rural Health Research and Policy Analysis Center.