

## MEASURING SMALL AREA VARIATION IN HOSPITAL USE: SITE-OF-CARE VERSUS PATIENT ORIGIN DATA

CATHERINE G. McLAUGHLIN

Department of Health Services Management and Policy, School of Public Health,  
The University of Michigan, MI 48109, U.S.A.

(Received for publication 21 March 1988)

**Abstract**—There has been increasing attention paid to small area variation in hospital discharge rates. While there is general agreement about the importance of correcting for the migration of patients to hospitals outside their geographic area when constructing population-based hospital use rates for these small areas, there have been no studies of the sensitivity of simple correlations or multiple regression results to these adjustments. Given the paucity of patient origin data, which is needed to adjust hospital discharge rates for patient crossovers, the problems of measurement error present in the more readily available site-of-care data need to be addressed. This paper analyzes the variation in hospital discharge rates, both an unadjusted site-of-care rate and an adjusted patient origin rate, across the 68 counties in the lower peninsula of Michigan in 1980. The results indicate that both simple correlations and multiple regression results of these rates with socio-economic and health care resource characteristics of the counties are very sensitive to the specification of the discharge rate, with the analysis of the unadjusted rate potentially leading to incorrect policy recommendations. The explanatory power of the socio-economic characteristics is underestimated and that of health care resource measures most likely overestimated when the discharge rate is not adjusted for patient crossovers.

### INTRODUCTION

Understanding why hospital use rates differ is important because of the need to forecast future utilization and in some cases, given the uncertainty about the desirability of such variations, to recommend appropriate policies to change high or low rates [1-3]. Although many studies have sought to explain the variation in hospital utilization rates across geographic areas, there is disagreement on the role of potential explanatory factors, such as socio-economic characteristics, availability of health care resources, and variations in physician practice styles. This is especially true for small area variation, which is of particular interest to most local and regional planners. Unfortunately, the data needed to analyze the determinants of small area variation in hospital rates are rarely obtainable.

There are three strategies that can be pursued by analysts interested in small area variation: (1) re-analyze existing, perhaps improved, data sets on the Medicare population and on the selected states for which patient origin data are available; (2) create new data sets, which is an expensive and time-consuming proposition; or (3) rely on site-of-care utilization data, which are readily available for all hospitals in the U.S. from the American Hospital Association's (AHA) Annual Survey. The purpose of this paper is to illustrate the problems associated with pursuing the third, and perhaps most important strategy by comparing the analysis of the variation in hospital discharge rates in Michigan calculated using the AHA site-of-care data to the analysis of the variation in hospital discharge rates using patient origin data.

### NATURE OF THE PROBLEM

Those interested in explaining the variation in hospital utilization and predicting future utilization are faced with two potential problems. The first is error in the measurement of utilization rates that results from relying on site-of-care data unadjusted for patient origin. The second is omitted variable bias, a methodological problem exacerbated by small samples or limited data on determining factors. While neither problem is unsolvable, obtaining the data needed to avoid both is difficult.

#### *Measurement error*

There is an important distinction between patient origin and site-of-care data. Patient origin data give hospital utilization information for the residents of an area, independent of the location of the hospital. Site-of-care data are utilization figures of hospitals, regardless of the residence of the patient.

One of the major advantages of patient origin data is the ability to develop a population-based data set. Utilization rates can be adjusted for patient "crossovers", either residents going to a hospital outside the area or nonresidents coming into area hospitals. In most cases, the rate is adjusted by changing the numerator to include all hospital utilization of residents in an area, regardless of the location of the hospital used by the residents. In contrast, site-of-care data do not make it possible to adjust for patient crossovers, and therefore result in error in the measurement of the utilization rate of a particular geographic area, underestimating the use rate of an area with a net outflow of residents to hospitals in other

areas and overestimating the use rate of an area with a net inflow.

When comparing the hospital utilization rate among geographic areas, it is important to correct for patient crossovers [4–6]. For example, Ginsburg and Koretz [4] point out that areas with higher than average bed-to-population ratios are more likely to have a large number of patients coming into the area to use those hospital facilities. Without correcting for this inflow of patients, the observed positive association between the bed rate in the community and the number of hospital admissions overstates the true relationship. Similarly, in a multiple regression, if the measurement error in the utilization rate is correlated with any included independent variables, the estimated coefficients on those variables will be biased.

#### *Omitted variable bias*

Many of the patient origin studies compare age-adjusted rates for various clinical procedures across small areas. These comparisons, and simple correlations between these rates and socio-economic and health system factors, have led some researchers to conclude that population characteristics are not correlated with use rates [7–11] and that some of the variance is explained by the availability of beds and physicians in the area [7–9]. These researchers also suggest that, to a large degree, the variation is due to differing physician practice styles.

A limitation of these population-based studies is the inability to control for multiple, or confounding, effects. Few of the studies using patient origin data have had a large enough sample and access to the data necessary to perform multiple regression analyses that control for these multiple effects. Although the few studies with controlled analyses have been limited to one segment of the population—either Medicare enrollees [4, 12, 13] or the residents of one state [14, 15]—and therefore may not yield generalizable results, all have shown that both socio-economic and health care resource factors are significant determinants of the variation in hospital utilization rates. Recently, Knickman and Foltz [16] found that population factors were more important in explaining the difference in use rates between New York City and Los Angeles, but that health care system characteristics were more important when comparing the use rate of 4 large Northeast Standard Metropolitan Statistical Areas (SMSAs) with 8 large Western SMSAs.

A major advantage of using site-of-care data is that they are usually available for many areas and are accompanied by data for other socio-economic and health care resource characteristics that may affect utilization. Studies using these data are therefore able to study hospital variation among a larger group of areas and to isolate the effects of particular population or health care services system characteristics through multiple regression techniques. Most of these studies have found that both socio-economic factors thought to influence demand, such as the age distribution of the population, and health system factors hypothesized to affect both demand and supply, such as the number of beds per capita, are statistically significant determinants of the variations in hospital use rates [17–20].

The results of the studies using multiple regression techniques indicate the need to control for the variation in both socio-economic and health care resource factors before concluding that the observed variation in utilization rates is due to some unobserved factor such as physician practice style, or to some observed, but not isolated, factor such as the supply of beds [1, 4, 6]. Failure to control for the contribution of these other factors in explaining the variation in use rates may result in omitted variable bias, and incorrectly concluding that the variable studied is the primary determinant, when it may in fact serve as a proxy for several omitted variables.

#### *Research needs*

Unfortunately, the data needs for a small area variation analysis of use rates adjusted for crossovers are often in conflict with the data needs of an analysis of use rates controlling for other determinants of hospital utilization. The first requires hospital utilization data with patient origin information for appropriately defined “small areas”; the second requires hospital utilization data for a large sample of areas for which there are complementary data bases.

To begin with, patient origin data are not readily available for most population groups. The samples of most national surveys, such as the Health Interview Survey, that do give utilization data by residence, are designed to yield national and regional estimates, not estimates of hospital use of small areas. These data cannot therefore be used to measure the variation in use across small areas. Theoretically insurance claims data exist for all groups, but in practice they are not accessible to most researchers. The one exception is the Medicare claims database, when can be used to examine the variation in hospital use by Medicare enrollees across small geographic areas. In addition, population-based data sets have been developed for several states, including 6 New England states [7, 8, 21], Washington [9], Manitoba [10, 11], Iowa [22], and Michigan [15].

The data requirements for a controlled analysis are quite stringent. Ideally, to answer the question of whether the factor of interest, for example bed supply, is responsible for the variation in hospital use rates, we need data on several areas that are identical in all factors that influence hospital use except for the number of beds per capita. Such data sets are rarely available. Some analysts have responded to the problem by looking at the change in utilization in one area over time, where they felt justified in assuming that only the factor of interest changed [23]. Others have matched areas according to a subset of use-determining factors, such as age and sex [11]. In both cases, if other, noncontrolled, factors are changing, the results can be misleading. To avoid this, a common approach is to employ a multiple regression technique in which an array of use-determining factors can be included and, therefore, the effect of a single factor can be isolated. A major limitation of this approach is the need to obtain data for all hypothesized determinants for a reasonably large sample, which presents a problem for many small area variation studies.

In summary, many of the patient origin studies have relied on simple correlations or rankings and are

therefore subject to omitted variable bias. The few patient origin studies that have been able to use multiple regressions and control for the many determinants of hospital use are limited to Medicare patients and a few geographic areas. AHA data, while allowing for a controlled analysis, do not have patient origin information and therefore will yield errors in the measurement of use rates. What follows is an illustration of the incorrect predictions and policy recommendations that could result from using these inappropriate, but accessible, data.

## AN ILLUSTRATION OF THE PROBLEM

### *The model*

The goal of the model of hospital utilization developed was to identify the population and health care system characteristics associated with the variation in hospital discharge rates across the 68 counties in the lower peninsula of Michigan in 1980, comparing the estimated effects of those characteristics on the variation in discharge rates using site-of-care data with their effects on discharge rates using patient origin data. A single-equation reduced-form regression model stemming from underlying behavioral models of supply and demand determination was estimated for each of the rates. Hospital use is specified as a function of exogenous factors that may determine the demand or the supply of hospital care. These factors fall into two general categories: population characteristics, such as socio-demographic and economic factors, and health care resource measures. The equation to be estimated takes the following general form:

$$D_i = a + bP_i + dH_i + e_i \quad (1)$$

where  $i$  denotes the unit of observation (in this case, county),  $D$  is the discharge rate,  $P$  is a vector of population characteristics,  $H$  is a vector of health care market factors, and  $e$  is the disturbance term.

### *Data*

The data needed to estimate this model came from two sources, the Area Resource File (ARF) and the Michigan Inpatient Data Base (MIDB). The ARF, compiled annually by the federal Department of Health and Human Services, is a county-based file consisting of data provided by a wide variety of sources, including the AHA, the American Medical Association (AMA), the National Center for Health Statistics (NCHS), and the Bureau of the Census. The data provided by the AHA are based on the annual survey of all hospitals in the U.S. The AMA data come from several surveys of member physicians, some of which are conducted annually, others which are repeated every few years. The data supplied by NCHS are similarly drawn from a variety of surveys, the frequency of which varies. The Census data are from the 1980 Census of Population and Housing. The MIDB is a patient origin database, capturing all hospital discharges for Michigan residents from short-term Michigan hospitals and from border hospitals in Indiana, Ohio, and Wisconsin.

### *Discharge rates*

Two different discharge rates were calculated and each was used as a dependent variable in the analysis.

The first discharge rate, based on AHA data, was calculated by dividing the total number of discharges from short-term general hospitals in county $_i$  by the population of county $_i$ . The resulting site-of-care discharge rate does not account for patient crossovers, that is, residents of county $_i$  admitted to hospitals in county $_j$  and residents of county $_j$  admitted to hospitals in county $_i$ . The derivation of this rate closely parallels that of several studies of the variation in hospital admission rates [17, 18, 20, 24, 25].

The second discharge rate, using MIDB data, follows the approach for adjusting for patient crossovers of Wennberg and Gittelsohn [7] and of most subsequent researchers who have relied on patient origin data. The rate was calculated by dividing the total number of discharges from any short-term general hospital in Michigan and nearby hospitals in Indiana, Ohio, and Wisconsin for the residents of county $_i$  by the population of county $_i$ . The resulting population-based rate captures the hospital use rate of the population in county $_i$ . Not all hospital discharges may be included, however. For example, it is possible that Michigan residents may enter a hospital in another state while on vacation, or choose to go to Mayo Clinic. Analysis of the Medicare claims database for 1980 reveals that a very small percent of hospital discharges for Michigan residents eligible for Medicare occur outside the state. The number of discharges outside the state for other population groups is not known. To the extent that these excluded discharges are not randomly distributed across counties in Michigan, but are correlated with other determinants of hospital use, the problems of measurement error still hold.

### *Explanatory variables*

The community characteristics included as explanatory variables in the model are described in this section (see also Table 1). As is always the case, there are several variables available to capture each hypothesized relationship, and various ways to measure each variable. Problems of multicollinearity and advantages of a parsimonious model make it undesirable to include every possible measure. The model was specified in light of the results of previous studies of the variation in hospital discharge rates. It is hypothesized in this model that each variable affects hospital use by determining either the demand or the supply of hospital services.

### *Population variables*

*Age: OLD, OLDER.* The need for and use of hospital services is hypothesized to increase as age increases. Most researchers have found that areas with a higher percent of the population over the age of 64 have higher admission rates [18, 20]. In addition, studies of Medicare enrollees also indicate that areas with proportionately more enrollees 75 and older, often referred to as the "old old," have higher admission rates than areas with a younger Medicare population [12, 13, 19].

*Economic factors: INCOME, POVFAM, BANK.* Several economic variables were included to capture both the individual's ability to purchase and the community's ability to provide hospital care, which are expected to lead to higher levels of hospital use

Table 1. Definitions, means, and standard deviations of variables: 68 counties in lower peninsula of Michigan, 1980

Variable	Definition; Mean (standard deviation)
<b>DEPENDENT VARIABLE: Discharge Rate</b>	
Site-of-care	Total number of discharges from short-term general hospitals in county, per 1,000 population in county, in 1980; 132.9 (85.2)
Patient origin	Total number of discharges from short-term general hospitals by residents of county, per 1,000 population in county, in 1980; 167 (30.3)
<b>INDEPENDENT VARIABLES:</b>	
<i>Population characteristics:</i>	
OLD	Percentage of population 65 to 75 years of age, 1980; 7.4 (2.3)
OLDER	Percentage of population 75 years of age and older, 1980; 4.4 (1.0)
POVFMFLY	Percentage of families below poverty line, 1979; 9.1 (3.0)
INCOME	Average personal income per capita, 1979; 7942 (1482)
COLLEGE	Percentage of those 25 years of age and older with a college education, 1980; 11.7 (5.4)
INFMORT	5-yr infant mortality rate per 1,000 births, 1974-78; 13.9 (2.4)
BANK	Total personal and corporate deposits in commercial banks and savings and loan institutions per capita, 1980; 4741 (1279)
Population size:	TINY = 1 if population was less than 25,000; 0.34 (0.48) SMALL = 1 if population was 25,000 to 50,000; 0.20 (0.41) MEDIUM = 1 if population was 50,000 to 250,000; 0.35 (0.48) LARGE = 1 if population was 250,000 or more; 0.10 (0.31)
<i>Health care resources</i>	
PHYSPC	Number of patient care physicians (MD and DO) per 1,000 population, 1980; 1.03 (0.86)
PHYSPR	Percentage of patient care physicians who are primary care physicians, 1980; 59.0 (18.9)

[15, 17]. However, to the degree that individual financial status also reflects the ability to purchase preventive care and to reduce the need for hospital services, income per capita will be negatively associated with hospital use and the percent of the families below the poverty line will be positively associated with use [18, 26]. While third-party coverage, both private and public, is hypothesized to affect hospital utilization, the only reliable information on coverage at the county level is the number of residents who are covered by Medicare. To the extent that private insurance coverage is correlated with income and that public coverage is correlated with age (Medicare) and poverty (Medicaid), by including these three variables in the model the effects of the related third-party coverage are also captured.

*Education: COLLEGE.* Education is hypothesized to be positively associated with health status and therefore negatively associated with hospital use [15, 26, 27].

*Mortality: INFMORT.* The infant mortality rate is intended to serve as a proxy for any differences in the overall health status of a community that are not captured by the other population variables. A standardized mortality rate adjusted for age and sex composition of the county population was also calculated and included in alternative regressions not reported here. The coefficients on the other variables did not change, and the mortality rate was not statistically significant.

*Population size: TINY, SMALL, MEDIUM.* In contrast to many of the states on which small area variation studies have been conducted that consist mainly of rural areas, Michigan is made up of rural, small city, and large urban areas. There are many reasons to expect the admission rate in rural areas to be different from that in small cities or urban areas. Population density has been hypothesized to reflect availability of care. Shorter commuting distances in heavily populated areas may make it easier to utilize ambulatory care. This substituting of ambulatory for inpatient care would lead to lower admission rates in more densely populated areas [17, 20]. Physician practice styles, particularly the propensity to admit, may differ in rural and urban areas [9]. Cultural attitudes toward health care and patterns of disease may also differ. To test for these differences, as well as to control for any omitted variables that are correlated with the size of the population, three population size dummy variables were added.

#### *Health care resource variables*

*Physician availability: PHYSPC, PHYSPR.* The direction of the effect of physician availability on hospital utilization is not clear. As pointed out by Eisenberg [30], there are many hypotheses regarding how and why physicians influence hospital utilization. If a higher concentration of physicians per capita indicates increased access to preventive and outpatient care, then hospital use in such an area would be lower. However, if a higher concentration causes individual physicians to face a decrease in patient load, then these physicians may respond by performing more procedures requiring hospitalization per patient in order to reach a targeted income. Again, empirical results differ, with studies of the Medicare population that used patient origin data supporting the hypothesized negative relationship [12, 13, 19], and studies of other population groups that used site-of-care data supporting the alternative hypothesis [17, 18, 24].

The predicted impact of having a greater proportion of physicians who are primary care physicians is more straightforward. Primary care physicians are more likely to treat patients outside the hospital. Previous research has consistently found that areas with a higher proportion of primary care physicians experience a lower hospital admission rate [15, 17, 18, 24].

*Hospital bed availability.* Another health care resource variable often included in analyses of the variation in hospital use rates is a measure of the availability of short-term hospital beds in the area. Following a study by Roemer and Shain [29], many researchers have hypothesized that a higher number of beds per capita in an area results in higher admission rates in that area. Several analysts have recognized the problem of endogeneity in this relationship—not only may higher bed availability lead to higher use rates, but higher use may also lead to higher bed rates [4, 18, 20]. Disentangling the two effects, while theoretically possible, is in practice very difficult. In addition, the bed rate in an area is itself some linear combination of most of the factors hypothesized to determine the use rate. When included as an explanatory variable, the bed rate often

dominates the individual effects of these other factors, thus making it difficult to assess their individual contributions to the variation in hospital use rates. For these reasons, bed rate is not included in this study. It is recognized that the estimated effects of certain included variables, such as physicians per capita, may be capturing not only the direct effect of physician supply on hospital use, but also the indirect effect of bed supply.

## RESULTS

### *Comparison of the two discharge rates*

Both the means and the standard deviations of the two county discharge rates, those using site-of-care data, not adjusted for patient crossover, and those using patient origin data, differ (Table 1). The mean value for the site-of-care discharge rates is considerably lower, 133 vs 167, reflecting the presence of 0 values. When using the AHA data, which give the number of discharges from member hospitals, the unadjusted discharge rate for the seven Michigan counties without hospitals is, by definition, equal to 0. This is clearly an underestimate of the true use rate of those counties. The problem is not only with the 0 values, however, but with virtually all of the observations. For example, those counties with large tertiary hospitals have very high unadjusted discharge rates, reflecting the inflow of patients from neighboring counties with smaller, less sophisticated community hospitals.

In addition, the error in measurement varies systematically with several community characteristics that have been found to affect hospital utilization. Counties with no hospitals or with only small hospitals (fewer than 100 beds) experienced a net outflow of residents and had significantly more older people, more poor families, lower levels of education, fewer physicians per capita, and a higher proportion of primary care physicians than did counties with at least one large hospital (greater than 300 beds), which experienced a net inflow of patients. Therefore, any simple comparison of community factors with these discharge rates will yield misleading conclusions, and any regression using the incorrect discharge rate will yield biased estimates of the effects of those variables.

An example of how such simple comparisons can be misleading is the relationship between the site-of-care discharge rates and the bed rates. Figure 1a shows a scatterplot of the unadjusted discharge rate and the ratio of short-term general beds per 1,000 population in each county. From this plot it would appear that there is a very strong positive relationship between the supply of beds and the discharge rate. Moreover, such a positive relationship is consistent with the results of several prior utilization studies. Once the discharge rates have been adjusted for patient crossovers, however, the relationship no longer holds, as can be seen in Fig. 1b, suggesting that hospitals in counties with high bed rates are used to a substantial extent by residents of other counties.

A correct analysis of the relationship between the availability of beds and use rates would adjust the bed rate to account for the proportion of beds available for use by nonresidents. Unfortunately there is no ready basis on which to allocate beds to

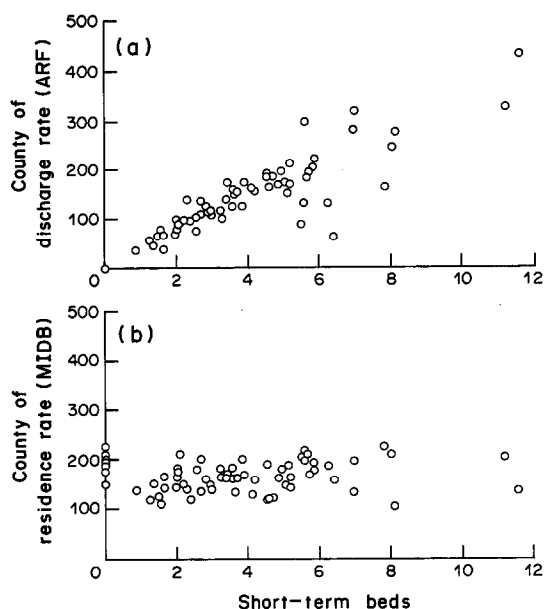


Fig. 1. (a) Unadjusted and (b) adjusted discharge rate vs bed rate.

nonresidents that is not itself a reflection of actual use, and therefore tautological. For that reason, apportioning beds according to use, a strategy often used, does not add to our understanding of the relationship between bed supply and use. For example, six of the seven Michigan counties with no hospitals have above average adjusted discharge rates. If beds from other counties are apportioned to these six counties according to use, then the adjusted bed rates for these counties will increase and reflect this higher use. The question of interest—why do residents of these counties with no hospital have higher than average hospital use rates?—remains unanswered, however.

Another interesting example is the relationship between use and education. As illustrated in Fig. 2a, the percentage of the county population 25 and over with a college education is positively associated with the county's unadjusted discharge rate. This positive relationship is contrary to expectations. It results from counties with large hospitals being more likely to have not only high unadjusted use rates but also a higher percentage of college graduates. In contrast, the negative relationship between education and use, which has been found in studies in which individuals were the unit of observation, is seen in Fig. 2b, when the adjusted discharge rate is used.

### *Regression results*

The results of the regression analyses, shown in Table 2, provide additional insight into the importance of correcting the discharge rate for crossovers, particularly when analyzing utilization at the county level. For both the adjusted and unadjusted hospital discharge rates, a fairly large percentage of the variation is explained by the model, although the model fares better with the unadjusted discharge rate, explaining 77% of the variation vs only 60% with the adjusted rate.

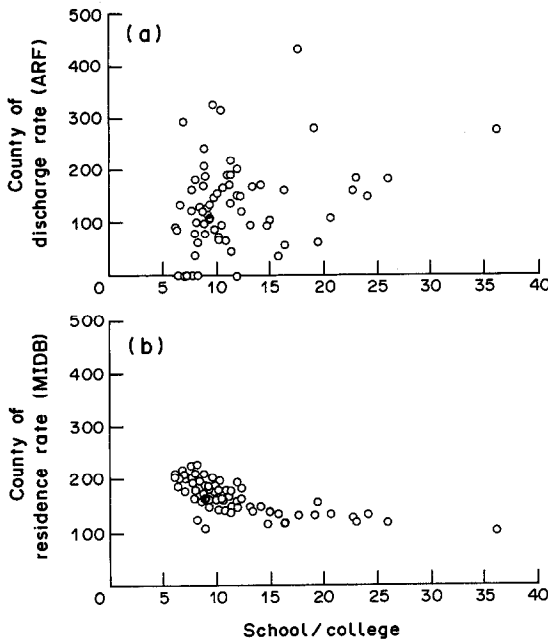


Fig. 2. (a) Unadjusted and (b) adjusted discharge rate vs college.

Another difference is in the relative contribution of different sets of variables. In the first set of regression, with the unadjusted site-of-care discharge rate as the dependent variable, population characteristics (OLD, OLDER, POVFAM, INFMORT, INCOME, COLLEGE) by themselves explain 18% of the variation in the discharge rate. Adding community

resource measures (BANK, TINY, SMALL, MEDIUM) explains an additional 28%. The health care resource measures (PHYSPC, PHYSPR) explain another 31% of the variation. These results imply that the availability of community and health care resources, measured by community wealth, size, and physician manpower, is the primary determinant of discharge rates.

The reverse pattern is true for the second set of regressions, using the adjusted patient origin discharge rate. In this case, the set of population characteristics explains 61% of the variation. Adding the community resource characteristics actually reduces the adjusted  $R^2$ . This suggests that when discharges are tied to the origin of the patient, the characteristics of the patient population are the important determinants of utilization, which is not a surprising result. With a discharge rate that is linked to the hospitals in a geographic area, the health care resource factors associated with the presence of hospitals, such as the number of physicians per capita and the percentage of those physicians who are primary care physicians, are going to be strongly correlated with use. The population based characteristics of county<sub>*i*</sub>, however, are being correlated with the use by *patients* in the hospitals in county<sub>*i*</sub>, not with the use of the *residents*, and are therefore not strongly associated with that use. When the discharge rate is corrected and then correlated with the characteristics of the population it represents, it is not surprising that these population characteristics as a group are significant determinants of that use.

A good illustration of the differences are the estimated effects of the variable OLD, the percentage of the population 65–74 years of age. With the un-

Table 2. Comparison of regression results, site-of-care discharge rates vs patient origin discharge rates: 68 lower peninsula Michigan counties, 1980

Variable	Unadjusted rate (AHA) (Site-of-care)			Adjusted rate (MIDB) (Patient Origin)		
CONST	-124.69 (115.96)	-95.02 (145.16)	257.76** (108.55)	125.99*** (28.61)	139.90*** (44.21)	155.75*** (50.79)
OLD	-31.94*** (10.74)	-31.77*** (9.99)	-23.15*** (6.65)	2.92 (2.65)	1.62 (3.04)	2.03 (3.11)
OLDER	47.92** (19.67)	16.87 (17.48)	0.89 (11.64)	1.37 (4.85)	0.18 (5.33)	-0.60 (5.45)
POVFAM	11.35** (5.35)	12.46*** (4.56)	6.76** (3.08)	2.38* (1.32)	2.22 (1.39)	1.93 (1.44)
INCOME	0.01 (0.01)	-0.006 (0.012)	-0.02** (0.01)	0.000 (0.002)	-0.001 (0.004)	-0.002 (0.004)
COLLEGE	3.41 (2.18)	3.30* (1.90)	-5.89*** (1.78)	-2.40*** (0.54)	-2.61*** (0.58)	-3.09*** (0.83)
INFMORT	3.98 (4.36)	2.47 (3.72)	3.94 (2.46)	1.16 (1.08)	1.17 (1.13)	1.25 (1.15)
BANK		0.04*** (0.01)	0.02*** (0.01)		0.003 (0.002)	0.002 (0.003)
TINY		73.99 (53.37)	112.82*** (35.87)		6.96 (16.26)	9.08 (16.78)
SMALL		71.47 (48.04)	107.90*** (32.74)		-0.07 (14.63)	2.01 (15.32)
MEDIUM		33.43 (36.19)	68.64*** (25.38)		-2.15 (11.02)	-0.12 (11.88)
PHYSPC			75.84*** (13.58)			4.11 (6.35)
PHYSPR			-1.89*** (0.42)			-0.08 (0.19)
$\bar{R}^2$	0.18	0.46	0.77	0.61	0.60	0.60

\*Significant at 0.90 confidence level.

\*\*Significant at 0.95 confidence level.

\*\*\*Significant at 0.99 confidence level.

adjusted discharge rate, the negative coefficient on OLD indicates that counties with an above average concentration of older people have below average discharge rates, a finding that contradicts expectations and the results of studies of individuals' hospital use. In actuality, this result is biased due to the error of measurement in the discharge rate, reflecting the fact that counties with no hospitals, or only small hospitals, and therefore underestimated discharge rates, tend to be characterized by older populations. Once the discharge rate is corrected for the use by these populations of large hospitals in other counties, the anticipated positive relationship is estimated.

The results also reflect the order in which the groups of variables were entered. The rationale for entering the population characteristics first is that physician supply factors are hypothesized to be determined in part by those same population characteristics. If entered first, the physician variables would capture the indirect effect of the population characteristics as well. Entering them last yields estimates of their net effect on discharge rates. In fact, if the health care resources variables are entered first, they explain 47% of the variation in the unadjusted rate, and 19% in the adjusted rate. The community resource variables then explain an additional 14 and 18% respectively, and the population characteristics, the remaining 16 and 23%. The relative significance of the health care resource factors increases in both cases, most noticeably for the adjusted discharge rates.

In the second set of regressions, where the adjusted patient origin rate is the dependent variable, all but one of the population variables do not have statistically significant coefficients. This may be attributable to multicollinearity between these factors, with several variables splitting the group's effect on the discharge rate. Only the level of education in the county appears to be statistically significant, even though the population characteristics as a group explain 61% of the variation.

It should also be noted that when analyzing the variation in the adjusted patient origin discharge rates, the unadjusted physician per capita rate may not capture the true effect of physician supply on hospital use. Measures of the physician in county<sub>*i*</sub> are appropriate to the extent that the physician in county<sub>*i*</sub> is the person making the decision either to admit the patient from county<sub>*i*</sub> to a hospital in county<sub>*i*</sub> or to refer a patient to a hospital in another county, which then results in an admission. The influence of physician availability in counties with major referral centers in determining the admission rates of patients from county<sub>*i*</sub> is ignored, however.

#### IMPLICATIONS FOR FUTURE RESEARCH

The results obtained in explaining small area variation in hospital use with the ARF unadjusted discharge data, the only data set for multiple population groups readily available to most researchers, are clearly misleading. However, without the comparison to the patient origin adjusted discharge data, many of the results would have appeared correct. For

example, the positive simple correlation between the site-of-care unadjusted discharge rate and the short-term bed rate is consistent with expectations. In fact, the lack of this strong positive relationship between the patient origin discharge rate and the bed rate is what causes concern and doubts over accuracy of measurement of the resource variables.

A relatively high adjusted  $R^2$  is obtained when regressing the unadjusted discharge rate on the independent variables and most of the estimated coefficients are statistically significant. With the exception of the variable OLD, the individual coefficients have the expected effect. The negative coefficient estimated for OLD could be interpreted as attributable to omitted variable bias or ecological fallacy. It is only by analyzing the distortions caused by patient crossovers that the reason for the negative coefficient is made clear.

Most multiple regression studies of the variation in hospital discharge rates, both those using patient origin data and those using site-of-care data, have found that both population and health care resource characteristics are important determinants of the variation in those rates. This is the first study to compare the estimates of the relative contribution of these two sets of characteristics in explaining the variation in site-of-care and in patient origin discharge rates. The results indicate that the explanatory power of the population characteristics and of health care resource measures differ according to the data analyzed, with the contribution of population characteristics underestimated and that of health care resource characteristics most likely overestimated when using unadjusted site-of-care discharge data. Forecasts based on such results and policy aimed at reducing the differences in discharge rates across geographic areas—whether to increase the lower rates or to decrease the higher rates—may erroneously focus on the supply of health care resources and ignore the role of various socio-economic characteristics in determining hospital utilization.

It is not clear whether the results of earlier studies of the variation in unadjusted site-of-care discharge rates suffer from the same problem. Most analysts who have used these discharge rates have studied states or SMSAs, not counties. Although the issue of patient crossover is not expected to be as serious a problem for a geographic area as large as a state or SMSA, the limited evidence available demonstrates that it cannot be dismissed. Figures recently published for Minnesota and Minneapolis–St Paul indicate that the admission rates adjusted for the inflow of patients for those two areas are lower than the unadjusted figures (in 1982, 145.7 vs 158.7 for the state and 128.7 vs 148.4 for Minneapolis–St Paul) [30]. Unfortunately, data on the outflow of patients were not available and therefore results reflecting those differences could not be obtained. Using Medicare claims data, Getts [31] reports that there is substantial migration by Medicare enrollees from one Health Systems Agency (HSA) area to another, with the proportion of discharges in an HSA attributable to residents in that HSA ranging from 0.6 to 1.8.

There are 12 SMSAs in Michigan. For five of the 12 SMSAs, the unadjusted site-of-care discharge rate was more than 15% higher than the adjusted patient

origin rate, in most cases reflecting an inflow of patients from neighboring rural areas. In one case, a small SMSA where the percentage of nonresidents using the hospitals in that SMSA was large relative to the population, the ranking of the SMSA changed from having the highest unadjusted discharge rate in the state to having the lowest adjusted rate. Similar switches in ranking for several HSA areas were reported by Getts for the unadjusted and adjusted discharge rates for Medicare enrollees [31]. Although such switching indicates a measurement problem, the real concern is whether the error in measurement is correlated with variables hypothesized to be determinants of hospital utilization. With only 12 observations, it was not possible to conduct regression comparisons at the SMSA level, but scatterplots showing the simple relationship between the discharge rates and several determinants reveals a pattern similar to that seen at the county level. For example, what appears to be a negative relationship between hospital use and the percentage of the population 75 yr and older using the unadjusted rate is actually due to measurement error, with the relationship between the adjusted discharge rate and OLDER positive as expected.

Given the paucity of patient origin data, local and regional planners will find it difficult to respond to the need for controlled analyses of small area variation. However, it is apparent from the results of this study that patient origin data are needed to accurately measure hospital utilization rates in small areas and, perhaps, even in not so small areas. Whether the same problems of mismeasurement and the resulting erroneous conclusions are present in analyzing the variation in hospital use rates among larger areas, such as SMSAs, is not known, but issues of measurement error must be addressed when forecasting future utilization in small geographic areas and formulating local policy aimed at affecting hospital utilization.

*Acknowledgement*—Research support for this paper was provided by the National Center for Health Services Research under grant No. HS05091.

## REFERENCES

1. F. D. Moore. Small area variations studies: illuminating or misleading? *Hlth Affairs* 4, 96–101 (1985).
2. M. R. Chassin, R. H. Brook, R. E. Park, J. Keesey, A. Fink, J. Kosecoff, K. Kahn, N. Merrick and D. H. Solomon. Variations in the use of medical and surgical services by the medicare population. *New Engl. J. Med.* 314(5), 285–290 (1986).
3. J. E. Wennberg. Which rate is right? *New Engl. J. Med.* 314(5), 310–311 (1986).
4. P. B. Ginsburg and D. M. Koretz. Bed availability and hospital utilization: estimates of the "Roemer effect". *Hlth Care Fin. Rev.* 5(1), 87–92 (1983).
5. J. E. Wennberg. Dealing with medical practice variations: a proposal for action. *Hlth Affairs* 3(2), 6–32 (1984).
6. S. A. Schroeder. Reviews: a medical educator. *Hlth Affairs* 3(2), 55–62 (1984).
7. J. E. Wennberg and A. Gittelsohn. Small area variations in health care delivery. *Science, N.Y.* 182, 1102–1108 (1973).
8. J. E. Wennberg and A. Gittelsohn. Variations in medical care among small areas. *Sci. Am.* 246(4), 120–134 (1982).
9. F. A. Connell, R. W. Day and J. P. LoGerfo. Hospitalization of Medicaid children: analysis of small area variations in admission rates. *Am. J. Publ. Hlth* 71(6), 606–613 (1981).
10. N. P. Roos and L. L. Roos. High and low surgical rates: risk factors for area residents. *Am. J. Publ. Hlth* 71(6), 591–600 (1981).
11. N. P. Roos and L. L. Roos. Surgical rate variations: do they reflect the health or socio-economic characteristics of the population? *Med. Care* 20, 945–958 (1982).
12. R. Deacon, J. Lubitz, M. Gornick and M. Newton. Analysis of variations in hospital use by medicare patients in PSRO areas, 1974–77. *Hlth Care Fin. Rev.* 1(1), 79–107 (1979).
13. P. A. Wilson. Hospital use by the aging population. *Inquiry* 18, 332–344 (1981).
14. J. Joffe. Mobility adjustments for small area utilization studies. *Inquiry* 16, 350–355 (1979).
15. P. Wilson and P. Tedeschi. Community correlates of hospital use. *Hlth Serv. Res.* 19(3), 333–355 (1984).
16. J. R. Knickman and A. Foltz. A statistical analysis of reasons for East–West differences in hospital use. *Inquiry* 22, 45–58 (1985).
17. M. S. Feldstein. Hospital cost inflation: a study of nonprofit price dynamics. *Am. Econ. Rev.* 61(5), 853–872 (1971).
18. B. R. Chiswick. Hospital utilization: an analysis of SMSA differences in occupancy rates, admission rates, and bed rates. *Explor. Econ. Res.* 3(3), 326–378 (1976).
19. M. Gornick. Trends and regional variations in hospital use under Medicare. *Hlth Care Fin. Rev.* 3(3), 41–73 (1982).
20. C. G. McLaughlin, J. C. Merrill and A. J. Freed. The impact of HMO growth on hospital costs and utilization. In *Advances in Health Economics and Health Services Research*, Vol. 5 (Edited by R. M. Scheffler and L. F. Rossiter). JAI Press, Washington DC. (1984).
21. B. A. Barnes, E. O'Brien, C. Comstock, D. G. D'Arpa and C. L. Donahue. Report on variation in rates of utilization of surgical services in the commonwealth of Massachusetts. *J. Am. Med. Association* 254(3), 371–375 (1985).
22. J. Wennberg. *Small Area Variations in Hospitalized Case Mix for DRGs in Maine, Massachusetts and Iowa*. National Center for Health Services Research, Washington D.C. (1984).
23. M. I. Roemer. Bed supply and hospital utilization; a natural experiment. *Hospitals* 35, 36–42 (1961).
24. K. Davis and L. B. Russell. The substitution of hospital outpatient care for inpatient care. *Rev. Econ. Statist.* 54(2), 109–120 (1972).
25. G. A. Melnick, J. R. C. Wheeler and P. J. Feldstein. Effects of rate regulation on selected components of hospital expenses. *Inquiry* 18(3), 240–246 (1981).
26. J. R. Knickman and A. Foltz. Regional differences in hospitalization: how much can be traced to population differences? *Med. Care* 22(11), 971–986 (1984).
27. M. Grossman. *The Demand for Health*. National Bureau of Economic Research, New York (1972).
28. J. Eisenberg. *Doctors' Decisions and the Cost of Medical Care*. Health Administration Press, Ann Arbor, MI (1986).
29. M. I. Roemer and M. Shain. *Hospital Utilization Under Insurance*. American Hospital Association, Chicago (1959).
30. B. E. Dowd. HMOs and twin cities admission rates. *Hlth Serv. Res.* 21(2), 177–188 (1986).
31. C. V. Getts. *Hospital Use: Health Status and Health Systems Report Series*. Institute for Health Planning, Madison, WI (1982).