

# A comparison of agency-based and self-report methods of measuring services across an urban environment by a drug-abusing homeless population

DAVID E. POLLIO,<sup>1,2</sup> CAROL S. NORTH,<sup>3</sup> KARIN M. EYRICH,<sup>4</sup> DOUGLAS A. FOSTER,<sup>2</sup>  
EDWARD L. SPITZNAGEL<sup>5</sup>

1 George Warren Brown School of Social Work, Washington University, USA

2 Department of Psychiatry, Washington University School of Medicine, USA

3 Department of Psychiatry, University of Texas Medical Center, USA

4 School of Social Administration, Temple University, USA

5 Department of Mathematics and Biostatistics, Washington University, USA

---

## Abstract

*The purpose of this paper is to advance the methodology for studying service assessment by comparing self-report and agency-generated methods. This study compares 30-day self-reported service use for homeless individuals (N = 229) randomly recruited from a single urban environment (St Louis, Missouri) with similar data collected from a broad array of service agencies providing homeless, substance abuse and outpatient mental health services across the same environment. Comparisons were made between self-report and agency-based data on shelter use, outpatient mental health service use (case management, psychiatric treatment, group therapy), outpatient substance abuse service use (case management, counselling, group therapy) and drop-in/day treatment use. Consistently low levels of kappa scores (all under 0.4) and correlation coefficients (only shelter use demonstrated significant agreement) were found. Findings demonstrated that the two methods of collecting service data are generally not concordant at the individual level. Certain demographic characteristics (increased age, being male, non-white ethnicity) and diagnoses (cocaine abuse/dependence, mania, schizophrenia) were associated with decreased reliability between the two methods of data collection. The two methods of assessment appeared to capture overlapping but not identical information. Each method of assessment has different utility to researchers and providers wishing to assess service use. Copyright © 2006 John Wiley & Sons, Ltd.*

**Key words:** services, service measurement, substance abuse, homelessness

---

In the literature on service utilization by the homeless population, type of services accessed (Pollio et al., 1997; Pollio et al., 2000; Koegel et al., 1999) and amount of services used (Barrow et al., 1987; Rife et al., 1991; Blankertz et al., 1992; Cohen et al., 1993; Morse et al., 1994; Schumacher et al., 1995; Pollio et al., 1997; Pollio et al., 2003) have both been found to predict the likelihood of achieving a variety of important outcomes.

One weakness of this research has been its lack of attention to methods of collecting service-use data. Two primary methods of data collection have been used in the literature: self-report by homeless respondents (Morse et al., 1994; Klinkenberg et al., 1998; Rosenheck et al., 1998; Weinreb et al., 1998; Chinman et al., 2000) and agency-based collections such as interviews with primary service providers (Dixon et al.,

1998; McGrew et al., 1999), reviews of agency records (Rife et al., 1991; Cox et al., 1998; Lam and Rosenheck, 1998; Humphreys and Rosenheck, 1998; Mallon, 1998), use of management information systems (Pollio et al., 2000), or a combination of self-report and agency-based information (Wolff et al., 1997).

The influence of methodological choices on findings, however, is not clear and may seriously affect the specific interpretation and application of these findings. Research examining agreement between self-report by homeless populations with mental illness and treatment staff estimates found reasonable agreement for most service categories (Calsyn et al., 1993). Other research, however, has questioned reporting rates and agreement between providers and homeless respondents on service use. Harrell (1997) found that use of drug treatment services in the past year was consistently under-reported by respondents compared to agency-derived records, suggesting that accuracy of self-reported service use may vary depending on the type of service estimated (Susser et al., 1989; Kashner et al., 1999).

Although the relationship between service use and reporting method is inadequately studied for homeless populations, literature on this topic for more general populations who abuse drugs gives clear direction for further inquiry. Studies of the drug-abusing population have found that accuracy of self-reported drug use varies by amount used and stigma associated with specific substances. For instance, studies have shown crack/cocaine use is under-reported in general (Morrall et al., 2000), as well as its use relative to other substances (Cook et al., 1997; Harrell, 1997; Magura and Kang, 1997; Wish et al., 1997). Young adults (Cook et al., 1997) and African-Americans (Falck et al., 1992; Febdrich and Vaugn, 1994; Page et al., 1997) have also been found to under-report substance use, in particular crack/cocaine. Thus, examining differential impacts of demographic, diagnostic, and substance use variables on reporting of service rates are a vital part of any methodological exploration.

The purpose of this study is therefore to explore differences and similarities further between self-report and agency-generated service use in the drug-abusing homeless population. Guided by previous research on homeless populations, as well as findings from methods studies of drug-using populations, the following research questions will be examined in this study:

1. Are there consistent differences in rates of specific services and types of services identified by self-report versus agency-generated methods?
2. Do specific demographic characteristics, diagnoses, and homeless subgroups consistently differ in rates by methods?

These questions generate two specific hypotheses:

- (a) in general, self-report methods will reveal bias towards under-reporting relative to agency-based data collection; and
- (b) individuals with diagnoses of substance abuse/dependence and severe mental illness will show less consistency in services received (both whether they received them or not and amount) between self-report and agency-based data collection than those without these diagnoses.

Specifically, this study will examine 30-day self-reported service use for homeless individuals randomly recruited from a single urban environment (St Louis, Missouri).

## Method

### Sampling

Subjects for this study constitute a portion of those recruited for a larger project (NIDA DA10713). Subjects were considered homeless if they (a) reported no current stable residence and (b) had spent the 14 previous nights staying in a public shelter, or in a park, an abandoned building, car, on the streets, or some other unsheltered location without a personal mailing address. Subjects who stayed with relatives or friends ('doubled up') were considered homeless if they had stayed in such a situation for more than six of the previous 14 nights. Subjects were also considered homeless if they had spent less than 30 days in an inexpensive hotel or motel and had no prior fixed address.

To capture a more representative homeless population than just the service-using segment of the homeless population (Smith et al., 1992, 1993), 80% of the sample was recruited randomly from 11 overnight shelters and one day shelter in St Louis and 20% from street locations across the city. Subjects were recruited from shelters in numbers proportionate to each shelter's census from a randomly generated computerized schedule, corresponding to that shelter's roster or bed count.

Recruitment for a given date and shelter ended when the designated one or two subjects were screened as eligible, indicated their willingness to participate, provided informed consent for participation in the study and scheduled a baseline interview. Subjects were considered eligible based on current homelessness and willingness/ability to provide additional information.

Street recruitment was conducted using 16 walking street routes, each with two or three identified starting points. Street routes were selected using known concentrations of homeless individuals within the city, as well as for geographic representation across the city. The order of routes traversed was randomized. The starting point from which recruitment was to begin along a given route was also randomized in order to equalize recruitment along each route over the course of the study. Interviewers administered the brief screener to all persons they encountered on a street route, not just those who 'appeared homeless'. Recruitment along a street route proceeded until a respondent screened in, signed an informed consent and scheduled a baseline interview. When no subject was found during the first time through a route, the same route was traversed again before moving on to the next route. If no subjects were found after two times through the second route, then street recruitment ended for the day.

The response rate among those who were eligible and willing to take part in the study was 94% (170/180) for shelter and 85% (42/49) for street recruitments, for an overall rate of 92% (212/229). Research by members of this team using similar recruitment strategies did not reveal significant differences between groups agreeing to participate and declining and found similar rates of refusal (for a more complete discussion see Smith et al., 1991, 1992, 1993).

#### *Instruments/measures*

Baseline data collection was accomplished for each subject by administration of a structured interview averaging three hours in length. Relevant sections of the interview for the current report included sociodemographic sections of the National Comorbidity Study (Kessler et al., 1994) interview; seven diagnosis sections of the Diagnostic Interview Schedule (DIS) (panic disorder, generalized anxiety disorder, mania, schizophrenia, antisocial personality disorder, depression, and post-traumatic stress disorder) (Robins et al., 1995); the residence section and selected questions from the DIS homeless supplement (North et al., 2004); a health

services section; and the alcohol, tobacco, and drug use sections of the Composite International Diagnostic Interview-Substance Abuse Module (Cottler and Compton, 1993).

In order to focus more specifically on reporting differences between methods rather than on recall by respondents, the data for self-reported service use in the 30 days prior to the interview were provided from the health services section of the baseline interview. When asking the questions in this section, interviewers produced a calendar to facilitate the respondent's accurate reporting of service use for the requested period of time. Although cognitive deficits cannot be ruled out, following recommendations by other researchers comparing reporting rates in a substance abusing population, choosing a shorter period of recall allows a more valid focus on reporting errors (Chermack et al., 2002).

Service use was solicited first as 'used' or 'not used' for the previous 30 days, without regard to length of use. For amount of use, number of service units were defined either using units familiar to and commonly used by both homeless and service provider (for example, shelter use as overnight stay) or by time period with conversion to units (for example, one unit of outpatient mental health or substance abuse is defined by the hour). Services operationalized by outside agencies (for example, in state mental health databases) used the agency's definition of self-report.

The agency-derived service use data for the same subjects were obtained in one of three ways, largely dependent upon the degree of sophistication of the target agencies. Data from most shelters and transitional housing providers in the city were extracted from an integrated management information system (MIS) designed and maintained for the city by a private firm with input from the parent project. Data from agencies not participating in the city's centralized intake, referral, and case-management network were obtained by manual review of agency rosters and sign-up sheets by project personnel. For state-funded community mental health centres, alcohol and drug treatment centres and psychiatric inpatient units in the city, service use data for study subjects were obtained from the state's Medicaid and Purchase of Service (POS) databases. The service categories used for comparison were shelter use, any outpatient mental health services, outpatient mental health case management, outpatient mental health psychiatric treatment (medication and diagnoses by psychiatrists), outpatient mental health group

therapy, any outpatient substance abuse services, outpatient substance abuse case management, outpatient substance abuse counselling/therapy, outpatient substance abuse group therapy and drop-in centre/day treatment.

The focus on these variables in the analysis is based on two factors – our confidence in the completeness of our data collection from both sources and our ability to match definitions exactly across all data sources. Because of these limitations, potentially informative service comparisons could not be made. For example, inpatient substance abuse treatment was included in the self-report methods but because inpatient records from all the major providers in the area were not available it was decided not to compare across reporting methods. Similarly, disparate definitions of counselling and therapy and psychiatric treatment across data collection sources precluded analysis of outpatient mental health services for counselling/therapy and outpatient substance abuse services for psychiatric treatment.

#### *Data analysis*

Descriptive and diagnostic data obtained for the sample are presented first. Service data were compared first at a bivariate level between self-report and agency-generated data, then through multivariate analyses.

Service data were coded dichotomously ('used'/'not used' during last 30 days) and continuously (number of units used). Dichotomous variables were compared using kappa coefficients. All service variables were included in the dichotomous analysis. Because the possibility existed that certain of the kappa scores reported were affected by the low base rate, we also examined the agreement for dichotomous variables using a Y statistic, following recommendations by Spitznagel and Helzer (1985). Because the results of this analysis do not raise scores for individual agreement up to acceptable levels (greater than 0.4), for ease of interpretation we have included kappa scores.

Continuous variables were compared using an estimate designed to compare correlation of a single fallible rater with a true state (Nunnally and Bernstein, 1994). For the current analysis this was obtained by computing the square root of the correlation coefficient of two fallible raters with each other. Note that the correlation coefficient is being used here to measure agreement rather than association. Thus, it serves much the same role as inter-class coefficient. Because of small percentages of reported service use across both methods, only

four service categories had sufficient distribution to be analysed – shelter use, any outpatient mental health service, any outpatient substance abuse service, and drop-in centre/day treatment. The continuous data were skewed, so a transformation was considered. However, because results were not substantially different between analyses of transformed and non-transformed, to ease interpretation non-transformed variables are reported in the results section.

Multivariate analysis focused on the four most frequently reported services – shelter use, any outpatient mental health services, any outpatient substance abuse services, and drop-in centre/day treatment, as the dependent variables. Logistic regression was used to predict simple agreement/disagreement between methods of service data collection. Differences in number of services reported were analysed using OLS regression. To control for skew and examine the absolute value of differences, differences were transformed using their square roots. This analytic strategy focused on size of difference rather than direction. This choice was made because large differences randomly distributed would cancel each other out, thus hiding substantial disagreements. The overall strategy was to generate simple regression (logistic/OLS) models for each independent variable, followed by a multiple regression including all significant variables in the final model. This data reduction strategy was chosen to address limitations imposed by the sample size and to ease interpretation of results. The results of this final analysis will be reported here.

Independent variables in the analysis include socio-demographic, diagnostic, and homelessness characteristics. Sociodemographic variables selected include age, race, and gender. Diagnostic variables selected include lifetime diagnoses (coded 'yes'/'no') for schizophrenia, mania, depression, adult antisocial personality disorder (ASPD), alcohol abuse/dependence, cocaine abuse/dependence, opiate abuse/dependence and a combined substance-abuse variable ('yes' for any abuse/dependence diagnosis versus 'no' for all three). Homelessness variables selected include age at onset of first homelessness and total length of time homeless.

## **Results**

### *Descriptive characteristics*

Table 1 presents descriptive statistics for the sample, which was primarily mature, male, and non-white, with

an average of 12 years of education. Mean age of onset of homelessness was  $33 \pm 14$  years, although the large standard deviation reflects a median age of 26 years. Subjects had been homeless an average of  $6 \pm 13$  years, although again the median was substantially lower (2 years).

Table 2 presents the percentages meeting lifetime and 12-month diagnoses for the sample. A strong majority of those sampled met criteria for at least a single substance abuse/dependence diagnosis lifetime (70%), with most of these also meeting criteria for the

previous 12 months (58% of the sample). The two most common single substance use disorders both for lifetime and 12 months were respectively alcohol (57% lifetime/39% 12 months) and cocaine abuse/dependence (44%/37%). Other diagnoses present in substantial proportions (at or greater than national prevalence estimates) included adult antisocial personality disorder, major depression, mania, and schizophrenia.

#### *Bivariate analysis/service use*

Considering the brevity of the service period, the sample might be characterized as consisting of relatively heavy service users. Table 3 presents service-use rates for the dichotomous variables for the two data collection methods with kappa scores. Approximately two-thirds of the sample (65%) self-reported using at least a single service type in the 30 previous days and more than four-fifths of the sample (84%) were reported in at least a single service type for the same period. Well over half had used shelters in the previous 30 days (by both report methods), slightly fewer than 1 in 5 had used drop-in centres or day-treatment programmes and more than 10% had used either mental health or substance-abuse treatment.

Overall percentages of use provided by the two data collection methods are superficially similar. Eight of the 10 categories were within 4% of each other. Of seven service types with disagreement between the two

**Table 1.** Descriptive characteristics

Characteristic	Percentage or mean (SD/SE)	N
Age in years	42 (11)	212
Percentage male	76 (2.9)	161/212
Percentage non-white	81 (2.7)	171/212
Years of education	11.9 (1.4)	212
Age of onset of homelessness in years	33 (14)	212
Average total lifetime years of homelessness	6 (13)	212

Sample N = 212

**Table 2.** Lifetime and 12-month diagnosis

Diagnosis	12 month			Lifetime		
	%	SE	N	%	SE	N
Substance abuse/dependence <sup>1</sup> (any alcohol or drug)	57	3.4	209	70	3.3	192
Drug abuse/dependence <sup>1</sup> (any cocaine or opiate)	37	3.3	209	46	3.7	181
Cocaine abuse/dependence <sup>1</sup>	36	3.3	209	43	3.6	186
Opiate abuse/dependence <sup>1</sup>	2	1.0	209	11	2.2	205
Alcohol abuse/dependence <sup>1</sup>	38	3.4	209	57	3.4	206
Adult antisocial personality disorder <sup>2</sup>	25	3.0	206	31	3.2	206
Major depression <sup>2</sup>	20	2.8	210	27	3.1	209
Mania (bipolar) <sup>2</sup>	9	2.2	165	12	2.5	163
Schizophrenia <sup>2</sup>	9	2.0	208	9	2.0	208

Note: diagnoses are calculated only for completed instruments. Differences in N reflect missing data.

<sup>1</sup>Diagnosis obtained using CIDI-SAM

<sup>2</sup>Diagnosis obtained using DIS

methods, self-report was higher in six. However, overall kappas were quite low – in the eight categories where kappas were generated, the overall average was 0.22 (range 0.12 to 0.39). Examination of the two-by-two tables used to calculate kappa coefficients revealed greater proportions in five of the six error cells (where one type reports service use and the other does not) of reported by self-report (SR) but not by agency-

generated (AG), than in the opposite diagonal (AG reporting and SR not reporting). This is consistent with the slightly higher overall rates of SR described previously.

Examination of correlations between the two types of service data collection again found the overall service rates superficially similar (presented in Table 4). On average, individuals used a substantial number of service

**Table 3.** Bivariate analysis of dichotomous service use for self-report (SR) versus agency-generated (AG) for 30 days: Percentage used, N, and Kappa scores

Service type	Percentage use (%)		N	Cell %		Kappa	CI	
	SR	AG		No/No Yes/No	No/Yes Yes/Yes		LL	UL
Shelter use	65	61	127	17 22	19 43	0.12	-0.05	0.30
Any outpt MH	12	10	137	82 8	7 4	0.25	0.02	0.49
Case Mgt	7	11	111	84 5	9 2	0.12	-0.12	0.37
Psychiatric tx	10	1	111	90 9	0 1	0.15	-0.11	0.42
Group therapy	3	0	109			—*		
Any outpt SA	11	11	137	82 7	7 4	0.33	0.09	0.57
Case mgt	8	8	108	87 5	5 4	0.39	0.09	0.70
Counseling/therapy	6	0	108			—*		
Group therapy	14	7	108	81 11	5 3	0.18	-0.07	0.43
Drop-in/Day tx	19	16	115	71 13	10 6	0.21	0.00	0.43

\*Kappa cannot be computed because of no reported use in one condition.

CI = Confidence Interval

LL = Lower limit

UL = Upper limit

**Table 4.** Bivariate analysis of continuous service use for self-report (SR) versus agency-generated (AG) for 30 days: mean units reported (SD), N, and corrected correlation coefficient

Service type	SR Mean	(SD)	AG Mean	(SD)	N	R	CI Lower	Upper
Shelter use	12.5	(14.0)	12.5	(14.7)	127	0.44*	0.14	0.60
Any outpatient MH	1.0	(4.7)	2.6	(12.6)	137	0.35	0.00	0.53
Any outpatient SA	1.2	(5.1)	4.0	(14.6)	137	0.22	0.00	0.46
Drop-in/Day treatment	3.9	(11.8)	2.4	(10.8)	124	0.14	0.00	0.44

\*p < 0.05

units within type, although the large standard deviations showed quite uneven distribution. Unlike the analysis of dichotomous service use reporting, no obvious biases were found between the two types of service use. Correlations between the two types of service use reporting were limited, averaging 0.29. One item – number of shelter stays reported – fell into a range generally considered acceptable. The two sources of information on shelter stays were significantly associated (unadjusted  $R = 0.20$ ,  $p = 0.03$ ).

#### *Multivariate analysis*

To clarify the impact of categories and characteristics on specific service types, this section's presentation will be organized by dependent variables, rather than by type of regression (logistic versus OLS). Similarly, to simplify interpretation, only final multiple regression models will be presented.

In the final logistic model predicting agreement between methods on shelter use during the time period, only two variables were included in the final equation – diagnosis of mania and cocaine abuse/dependence. The equation as a whole was significant (model chi-square = 7.32,  $df = 2$ ,  $n = 94$ ,  $p = 0.02$ ). Of the two diagnoses in the final model, only mania was significant. Persons meeting diagnostic criteria for mania increased odds of disagreement by a factor of 5.3 between the two types of methods of data collection (confidence interval (CI) 0.97 – 27.83; chi-square = 3.82,  $p = 0.05$ ). Although the final OLS model was significant as a whole ( $F_{1,86} = 4.69$ ,  $p = 0.03$ ), only a single variable, age, was included. One year of increased age was associated with an increased difference of  $0.05 + 0.02$  units in reported number of stays ( $t = 2.16$ ,  $p = 0.03$ ).

The final logistic model predicting agreement on outpatient mental health service use was significant as a whole (model chi-square = 13.9,  $df = 4$ ,  $n = 129$ ,  $p = 0.008$ ) and included four variables, three of which were significant in the final analysis. Non-white ethnicity was associated with increased odds of disagreement by a factor of 4.4 between methods (CI 1.02 – 11.11; chi-square = 5.41,  $p = 0.02$ ); male gender increased odds of disagreement by a factor of 4.4. (CI 1.08 – 10.23; chi-square = 4.37,  $p = 0.04$ ); and persons with schizophrenia increased odds of disagreement by a factor of 4.1 (CI 1.09 – 15.68; chi-square = 4.36,  $p = 0.04$ ). Diagnosis of opiate abuse/dependence was not significant in the final analysis. For OLS models, the equation as a whole

was significant ( $F_{4,112} = 9.55$ ,  $p > 0.0001$ ), with four variables in the final model, three of which independently proved significant. Male gender increased difference by  $1.2 + 0.6$  ( $t = 4.50$ ,  $p < 0.0001$ ), each year of age increased difference  $0.02 + 0.01$  hours ( $t = 2.06$ ,  $p = 0.05$ ), and diagnosis of schizophrenia increased difference  $0.8 + 0.3$  hours ( $t = 2.63$ ,  $p = 0.01$ ) in service units reported between methods. Diagnosis of major depression was not significant in the final model.

The final logistic model predicting agreement on reported use of outpatient substance-abuse services was significant as a whole (model chi-square = 41.04,  $df = 3$ ,  $n = 129$ ,  $p < 0.0001$ ) and included three variables, two of which were significant in the final analysis. Male gender increased odds of disagreement between methods by a factor of 32.3 (CI 7.22 – 144.46; chi-square = 20.68,  $p < 0.0001$ ) and cocaine abuse/dependence diagnosis increased odds of disagreement by a factor of 9.5 (CI 2.27 – 40.12; chi-square = 9.46,  $p = 0.001$ ). Diagnosis of schizophrenia was not significant in the final model. The final OLS model was quite similar to the logistic model. The equation as a whole was significant ( $F_{3,113} = 14.72$ ,  $p > 0.0001$ ), with male gender increasing differences  $1.4 \pm 0.2$  hours ( $t = 5.83$ ,  $p > 0.0001$ ) and cocaine abuse/dependence by  $0.5 \pm 0.2$  hours ( $t = 2.52$ ,  $p = 0.01$ ) in service use reported between methods. Although the diagnosis of schizophrenia was included in the final equation, it was not significant in the final model.

In the final set of analyses, the logistic regression predicting disagreement in service use between types as a whole was significant (model chi-square = 5.13,  $df = 1$ ,  $n = 115$ ,  $p = 0.02$ ) but the model included only a single variable, non-white ethnicity, increased odds of disagreement between methods by a factor of 4.4 (CI 1.22 – 10.1 chi-square = 5.38,  $p = 0.02$ ). For the OLS regression, no variables were significant predictors of differences.

#### **Discussion**

The first research question was: 'Are there consistent differences in rates of specific services and types of services identified by self-report versus agency-generated methods?' At the superficial level of aggregate rates, the initial answer appears to be that the two methods provided very similar overall rates of service use. However, in examining the reliability of these responses for service categories at the individual reporting level, the consistently low kappa scores and correla-

tion coefficients argue strongly that the two methods of collecting service data are not reliable at the individual level. Error patterns appear to go consistently in both directions.

Thus, the results fail to support our first hypothesis predicting general under-reporting by self-report methods. A possible interpretation more in line with the general findings might be that the two data collection methods are actually collecting conceptually related but not identical information. It may be that self-report of services is guided by a set of principles that do not correspond to the reality of a given time frame. One possible explanation may be that individuals identify only the personally 'useful' services (that is, services that were personally important), rather than actual ones. Another explanation could be that there are painful memories associated with specific service use, decreasing the likelihood of service being recalled. Thus, despite efforts to constrain self-report to a consistent time frame, results may be biased towards over-reporting these figural services outside the time period.

Similarly, agencies may be reporting only services that are congruent with their mission. For example, a shelter may provide opportunities for informal group meetings between night staff and homeless persons but it may simply report providing a bed, while the homeless person may recall the night as including both a bed and a group – a bias that an MIS-based approach may not adequately capture. Although both of these interpretations represent conjecture beyond the data themselves, the explanation represents a strong argument against an assumption of one type of service data collection as representing a 'gold standard'.

Another potential impact on the lack of congruence between the two methods may lie within characteristics of the organizations themselves. Although it is not possible to test this given the current data, we have noted the differences in levels of sophistication and precision of data collection from participant agencies. A possibility (and potential future research direction) would be to examine whether differences within organizations (for example, available technology, number of different services offered within the agency, organizational culture, importance of accurate data collection) may impact on agreement rates.

The second research question was: 'Are there consistent differences in these rates by methods for specific demographic characteristics, diagnoses, and homeless

sub-groups?' In general, although specific variables provided some findings, relatively few significant relationships were identified relative to the number of analyses. In particular, it was noteworthy that no homeless variables were significant in any of the equations.

Table 5 presents a summary of the significant multivariate results organized by independent variables (specific odds ratios, confidence intervals, etc. are reported in the multivariate analysis part of the results section). This figure structures the findings to orient the examination to the second hypothesis: that specific demographic characteristics, substance abuse/dependence diagnoses and severe mental illness will generally manifest decreased consistency of self-report relative to agency-based data collection. Overall, impact of diagnosis on reporting reliability may be an important factor for services that are congruent with specific psychiatric diagnoses (here cocaine abuse/dependence, mania, and schizophrenia) and service types (substance abuse and mental health treatment, and shelter use), rather than acting as a general phenomenon. Individuals with a cocaine-abuse/dependence diagnosis had greater likelihood to disagree between methods and greater differences in reported number of services. This finding follows the argument reported in the literature that stigma associated with cocaine increases likelihood of an individual denying use (Susser et al., 1989, 1990), as well as the previously discussed research for non-homeless populations (Cook et al., 1997; Harrell, 1997; Magura and Kang, 1997; Morral et al., 2000).

Among demographic characteristics, greater age, male gender, and non-white ethnicity were all associated with decreased reliability, male gender showing the most consistency in this regard. Although the current research cannot specifically inform the various mechanisms for lack of reliability for demographic characteristics, these characteristics may reflect a greater impact of stigma on reporting type for specific demographic categories.

The findings of this study present important information for research and for assessment for service needs. Because lack of a pattern of consistent finding does not allow a general conclusion about how the two types of data collection methods are related (whether one under-reports relative to the other), researchers must be aware of potential biases introduced by specific individual characteristics of each method. We argue that neither represents a 'gold standard' methodologically, so the conclusion must be that either both must



**Table 5.** Summary of significant multivariate results organized by independent variables

Independent predictor	Dependent variable (services by sector)	Direction of difference	p value
<b>Diagnoses</b>			
Cocaine	Substance abuse services	Greater difference	p = 0.01
		Disagree	p = 0.001
Mania	Shelter use services	Disagree	p = 0.05
Schizophrenia	Mental health services	Disagree	p = 0.04
		Greater difference	p = 0.01
<b>Demographics</b>			
Age (younger)	Shelter use services	Greater difference	p = 0.03
	Mental health services	Greater difference	p = 0.05
Male	Mental health services	Greater difference	p < 0.0001
		Disagree	p = 0.04
	Substance use services	Greater difference	p < 0.0001
		Disagree	p < 0.0001
Non-white	Mental health services	Disagree	p = 0.02
	Drop-in/Day tx services	Disagree	p = 0.02

be collected or the choice of method must be based on the specific question asked. For example, cost analyses would be likely to be better estimated using agency records. In terms of service use, agency assessment is generally conducted in a lower-resource environment so it is likely that the relative simplicity of asking the individual for current service use may be deemed acceptable – not necessarily ‘the truth’ but able to represent sufficiently useful information.

#### *Limitations*

As an exploration of this issue for the homeless population, this study has significant limitations. Perhaps most importantly, the comparison was between service use for a relatively brief period of time (30 days). Although this methodological choice enhances the reliability of recall, it also may introduce biases, particularly for the dichotomous ‘used’/‘not used’ comparisons. Certainly, future research needs to explore why and how homeless persons recall the services that they have used.

A second limitation of the study lies in the impact of missing data on analytic sample size. A variety of factors (incomplete data from subjects, inconclusive diagnoses derived from instrumentation, missing agency data) impacted on availability of data. Thus, results must be considered with some caution.

A third limitation in comparing methods is the degree of completeness of agency-generated records versus recall. Although the agencies participating in

the current study represented nearly the entire set of major providers of the service types within the metropolitan area, they hardly represented all of the possible providers available to study subjects. For example, self-report might include services from agencies outside of the metropolitan St Louis area, which agency-generated reports could not capture. Additionally, the definition of outpatient mental health services allowed the possible inclusion in self-report of private therapy or religious counselling, two interventions not reached by agency-generated data. Thus, additional error is likely to be introduced into this analysis based on this limitation.

A fourth limitation concerns the measurement of service units. For each type, it was assumed that self-report and agency-generated service data measured service unit identically. Obviously, this represented an unwarranted assumption about the precision of measurement. More precise measurement of services could perhaps have yielded somewhat different results. Yet another untested assumption of the current study is that MIS-generated data and agency-records yield identical data. Finally, many additional variables were under-specified. The limited information on service use, interviewer effects, homelessness, demographic variables, and current state of cognitive impairment (through level of symptoms, current intoxication, or objective testing) might all contribute additional insight to subsequent research.

## Acknowledgements

This research was supported by NIDA grant DA10713 to Dr North. The authors wish to acknowledge the support of the research project staff.

## References

- Barrow SM, Hellman F, Lovell AM, Plapinger JD, Struening EL. Effectiveness of Programs for the Mentally Ill Homeless. New York: New York Psychiatric Institute, 1989.
- Blankertz LE, Cnaan RA, Saunders M. Assessing the impact of serving the long-term mentally disabled homeless. *Journal of Sociology and Social Welfare* 1992; 19: 199–220.
- Calsyn RJ, Allen G, Morse GA, Ruth, S, Templehoff B. Can you trust self-report data provided by homeless mentally ill individuals? *Eval Rev* 1993; 17(3): 353–66.
- Chermack ST, Roll J, Reilly M, Davis L, Kilaru U, Grabowski J. Comparison of patient self-reports and urinalysis results obtained under naturalistic treatment conditions. *Drug Alcohol Depend* 2002; 59: 43–9.
- Chinman MJ, Rosenheck R, Lam JA. The case management relationship and outcomes of homeless persons with serious mental illness. *Psychiatr Serv* 2000; 51(9): 1142–7.
- Cohen C, Onserud H, Monaco C. Outcomes for the mentally ill in a program for older homeless persons. *Hospital and Community Psychiatry* 1993; 44: 650–6.
- Cook RF, Bernstein AD, Andrews CM. Assessing drug use in the workplace: A comparison of self-report, urinalysis, and hair analysis. In L Harrison, A Hughes (eds) *The Validity of Self-Reported Drug Use: Improving the Accuracy of Survey Estimates*. NIDA Research Monograph #167. Washington D.C.: National Institute on Drug Abuse, 1997.
- Cottler LB, Compton WC. Advantages of the CIDI family of instruments in epidemiological research on substance use disorders. *Int J Meth Psychiatr Res* 1993; 3: 109–19.
- Cox GB, Walker RD, Freng SA, Short BA, Meijer L, Gilchrist L. Outcome of a controlled trial of the effectiveness of intensive case management for chronic public inebriates. *J Stud Alcohol* 1998; (September): 523–32.
- Dixon L, Stewart B, Krauss N, Robbins J, Hackman A, Lehman A. The participation of families of homeless persons with severe mental illness in an outreach intervention. *Community Ment Health J* 1998; 34(3): 251–9.
- Falck R, Siegel HA, Forney MA, Wang J, Carlson RG. The validity of injection drug users' self-reported use of opiates and cocaine. *J Drug Issues* 1992; 22(4): 823–32.
- Fendrich M, Vaughn CM. Diminished lifetime substance use over time: an inquiry into differential under-reporting. *Publ Opin Q* 1994; 58: 96–123.
- Harrell AV. The validity of self-reported drug use data: the accuracy of responses on confidential self-administered answered sheets. In L Harrison, A Hughes (eds) *The Validity of Self-Reported Drug Use: Improving the Accuracy of Survey Estimates*. NIDA Research Monograph #167. Washington D.C.: National Institute on Drug Abuse, 1997, pp. 37–58.
- Humphreys K, Rosenheck R. Treatment involvement and outcomes for four subtypes of homeless veterans. *Am J Orthopsychiatry* 1998; 68(2): 285–94.
- Kashner TM, Suppes T, Rush AJ, Altshuler KZ. Measuring use of outpatient care among mentally ill individuals: a comparison of self-reports and provider records. *Evaluation and Program Planning* 1999; 22: 31–9.
- Kessler RG, McGonagle KA, Zhao S, Nelson CB, Hughes M, Eshelman S, Wittchen HU, Kendler KS. Lifetime and 12-month prevalence of DSM-III-R psychiatric disorders in the United States: Results from the National Comorbidity Study. *Arch Gen Psychiatr* 1994; 51: 8–19.
- Klinkenberg WD, Calsyn RJ, Morse GA. The helping alliance in case management for homeless persons with severe mental illness. *Community Ment Health J* 1998; 34(6): 569–78.
- Koegel P, Sullivan G, Burnam A, Wenzel S. Utilization of mental health and substance abuse services among homeless adults in Los Angeles. *Med Care* 1999; 37: 306–17.
- Lam JA, Rosenheck R. The effect of victimization on clinical outcomes of homeless persons with serious mental illness. *Psychiatr Serv* 1998; 49(5): 678–83.
- Magura S, Kang S. The validity of self-reported cocaine use in two high-risk populations. In L Harrison, A Hughes (eds) *The Validity of Self-Reported Drug Use: Improving the Accuracy of Survey Estimates*. NIDA Research Monograph #167. Washington D.C.: National Institute on Drug Abuse, 1997, pp. 227–46.
- Mallon GP. After care, then where? Outcomes of an independent living program. *Child Welfare* 1998; 77(1): 61–78.
- McGrew JH, Wright ER, Pescosolido BA, McDonel EC. The closing of central state hospital: Long-term outcomes for persons with severe mental illness. *The Journal of Behavioral Health Services and Research* 1999; 26(3): 246–61.
- Morrall AR, McCaffrey D, Iguchi MY. Hardcore drug users claim to be occasional users: Drug use frequency under-reporting. *Drug Alcohol Depend* 2000; 57: 193–202.
- Morse GA, Calsyn RJ, Allen G, Kenny DA. Helping homeless mentally ill people: what variables mediate and moderate program effects? *Am J Community Psychol* 1994; 22(5): 661–83.
- North CS, Eyrich KM, Pollio DE, Foster DA, Cottler LB, Spitznagel EL. The homeless supplement to the Diagnostic Interview Schedule: test-retest analysis. *Int J Meth Psychiatr Res* 2004; 13(3): 184–90.
- Nunnally JC, Bernstein IH. *Psychometric Theory*. 3rd edn. New York: McGraw-Hill, 1994.
- Page WF, Davies JE, Ladner RA, Alfasso J, Tennis H. Urinalysis screened versus verbally reported drug use: The identification of discrepant groups. *Int J Addiction* 1977; 12(4): 439–50.

- Pollio DE, North CS, Eyrich KM, Foster DA, Spitznagel EL. Modeling service access in a homeless population. *J Psychoactive Drugs* 2003; 35(4): 487–96.
- Pollio DE, North CS, Thompson SJ, Paquin JW, Spitznagel EL. Predictors of achieving stable housing in a mentally ill homeless population. *Psychiatr Serv* 1997; 44(4): 528–30.
- Pollio DE, Spitznagel EL, North CS, Thompson SJ, Foster DA. Service use over time and achievement of stable housing in a mentally ill homeless population. *Psychiatr Serv* 2000; 51(12): 1536–43.
- Rife JC, First RJ, Greenlee RW, Miller LD, Feichter MA. Case management with homeless mentally ill people. *Health and Social Work* 1991; 16(1): 58–67.
- Robins L, Cottler L, Bucholz K, Compton W. Diagnostic Interview Schedule for DSM-IV. St Louis MO: Washington University, 1995.
- Rosenheck R, Morrissey J, Lam J, Calloway M, Johnsen M, Goldman H, Randolph F, Blasinsky M, Fontana A, Calsyn R, Teague G. Service system integration, access to services, and housing outcomes in a program for homeless persons with severe mental illness. *Am J Publ Health* 1998; 88(11): 1610–15.
- Schumacher JE, Milby JB, Caldwell E, Raczynski J, Engle M, Michael M, Carr J. Treatment outcome as a function treatment attendance with homeless persons abusing cocaine. *J Addict Dis* 1995; 14(4): 73–85.
- Smith EM, North CS, Spitznagel EL. Are hard-to-interview street dwellers needed in assessing psychiatric disorders in homeless men? *Int J Meth Psychiatr Res* 1991; 1: 69–78.
- Smith EM, North CS, Spitznagel EL. A systematic study of mental illness, substance abuse and treatment in 600 homeless men. *Ann Clin Psychiatr* 1992; 4: 111–20.
- Smith EM, North CS, Spitznagel EL. Alcohol, drugs, and psychiatric comorbidity among homeless women: an epidemiologic study. *J Clin Psychiatr* 1993; 54: 82–7.
- Spitznagel EL, Helzer JE. A proposed solution to the base rate problem in the kappa statistic. *Arch Gen Psychiat* 1985; 42: 725–8.
- Susser E, Conover S, Struening EL. Problems of epidemiologic method in assessing the type and extent of mental illness among homeless adults. *Hospital and Community Psychiatry* 1989; 40: 261–5.
- Susser E, Conover S, Struening EL. Mental illness in the homeless: problems of epidemiologic method in surveys of the 1980s. *Community Ment Health J* 1990; 26: 391–412.
- Weinreb L, Goldberg R, Bassuk E, Perloff J. Determinants of health and service use patterns in homeless and low-income housed children. *Pediatrics* 1998; 102(3): 554–62.
- Wish ED, Hoffman JA, Nemes S. The validity of self-reports of drug use at treatment admission and at follow-up: Comparisons with urinalysis and hair assays. In L Harrison, A Hughes (eds) *The Validity of Self-Reported Drug Use: Improving the Accuracy of Survey Estimates*. NIDA Research Monograph #167. Washington D.C.: National Institute on Drug Abuse, 1997, pp. 200–26.
- Wolff N, Helminiak TW, Morse GA, Calsyn RJ, Klinkenberg WD, Trusty ML. Cost-effectiveness evaluation of three approaches to case management for homeless mentally ill clients. *Am J Psychiatr* 1997; 154(3): 341–8.

*Correspondence: David E. Pollio, Associate Professor of Psychiatry and Social Work Washington University, Campus Box 1196, St Louis, MO 63130-4899, USA. Telephone (+1) 314 935-7516. Fax (+) 314 935-8511. Email: depollio@wustl.edu.*