

Does Multisite Sampling Improve Patient Heterogeneity in Drug Misuse Research?

John Dunn^{1*}, Cleusa Pinheiro Ferri², Ronaldo Laranjeira²

¹Academic Department of Psychiatry, Royal Free and University College Medical School, Royal Free Hospital, Pond Street, London, NW3 2QG, U.K.

² UNIAD (Unidade de Pesquisa em Alcool e Drogas), Departamento de Psiquiatria, Escola Paulista de Medicina, Universidade Federal de São Paulo, Rua Botucatu, 740, São Paulo-SP, 04023-900, Brazil

* Corresponding author. Tel: 00 44 20 7794 0500 ext 3719/3950. E-mail: drjdunn@hotmail.com

Abstract

The aim of this study was to investigate whether multisite sampling increased heterogeneity among a sample of cocaine users from São Paulo, Brazil. Six hundred and fourteen cocaine users were interviewed at 23 fixed sites plus an out-of-treatment sample. The sites were then re-grouped into six main types: university outpatient clinics, public outpatient clinics, public inpatient units, private inpatient units, HIV services and non-treatment. Marked differences were found between users recruited at these sites, especially in relation to age, gender, employment status, criminal history, history of prostitution, previous drug misuse treatment, duration of cocaine use and lifetime use of intravenous cocaine. These results suggest that multisite sampling is a valid method for increasing patient heterogeneity, but whether it improves representativeness and thus the generalisability of drug misuse research is debatable.

Keywords: Substance Misuse, Cocaine, Epidemiological Methods, Sampling, Brazil

1. Introduction

Probability sampling is the gold standard for obtaining representative samples of individuals from a population. However, when drug misusers are the target group various problems arise which mean that probability sampling can become impracticable (Dunn & Ferri, 1999). First, there is the problem of low prevalence. For example, an estimate from the UK suggests that around 2% of the general population are dependent on illicit drugs (Farrell et al, 1998). However, it is likely that less than a quarter of these are dependent on individual substances, such as cocaine. Consequently, to find just one cocaine addict, one might need to interview at least 400 people. To find enough cocaine addicts (300 for example) to allow us to make statistically powerful statements about their characteristics, we might need to interview over 120,000 people. The larger the sample, the longer it will take to complete the interviews, the more expensive the study becomes and since the ratio of cases to non-case is very low, the whole process becomes extremely inefficient.

A second problem with probability sampling is that illegal activities, such as drug misuse, may be hidden or denied, so it can be difficult to find or identify users. Surveys tend to be based on the occupants of private households, so cocaine addicts may be missed because they are out buying drugs or involved in illegal activities to finance their use or because they are homeless, in prison or in residential treatment (Dunn & Ferri, 1999). Consequently many research studies that aim to investigate the characteristics of drug users and their patterns of use tend to employ convenience samples using patients from outpatient and inpatient units (Griffin et al, 1989; Kleinman et al, 1990; Dunn et al, 1996). However, selection bias means that drug users recruited from these settings have very different characteristics to those who are not in treatment. Studies suggest that drug users not in treatment are more likely to be polydrug users, be involved in illegal

activities but have fewer negative consequences of drug use and less depression (Rounsaville & Kleber, 1985; Carroll & Rounsaville, 1992).

If only one treatment service were used for sampling then the patients accrued might be very homogeneous due to factors such as geographic location, public or private status, type of treatment offered and admission criteria. In recent years attempts have been made to recruit more patients from non-treatment settings using snowball sampling (Lopes et al, 1996), privileged access interviewer techniques (Griffiths et al, 1993) or by directly approaching suspected drug users on the streets (Edlin et al, 1994), but none of these methods is immune from selection bias. Patients nominating drug-using friends or colleagues who are not in treatment, may be more likely to indicate people who share similar sociodemographic and drug use characteristics as themselves.

Multisite sampling has been offered as a way of increasing patient heterogeneity and representativeness (Haw et al, 1992). With this method, although sampling bias may still operate at each site, it is postulated that with a large enough number of sites, these biases will tend to cancel each other out. However, we cannot assume that this will happen, if the biases are mainly acting in the same direction, then the final sample may not be as heterogeneous or representative as was hoped. Multisite sampling is widely used in research studies of drug misuse and HIV (Lewis et al 1992; Gossop et al, 1998; Deren et al, 1998) but is it valid? The aim of this study was to investigate whether multisite sampling increased heterogeneity among a sample of cocaine users.

2. Methods

The sample used in this analysis was combined from two separate studies that had been undertaken concurrently. The first sample (*Study I*) was from Dunn and Laranjeira's (1999a) study of the profiles, patterns of drug use and HIV-risk behaviours

among cocaine users in Sao Paulo. The second sample (*Study II*) came from Ferri's and Gossop's (1999) study comparing treatment seeking and out-of-treatment cocaine users, also in Sao Paulo. Although each study had different objectives, their questionnaires had many identical items, in particular those concerning demographic details and drug misuse history. More detailed information about the methodology used in each study can be obtained from previously published work, in particular Dunn & Laranjeira (1999b) and Ferri & Gossop (1999).

2.1 Setting

This was a cross sectional study with drug users interviewed between January 1996 and December 1997. In total 23 different treatment and counselling services were used in addition to an out-of-treatment sample. The characteristics of each site are summarised in Table 1. All services are situated in the State of Sao Paulo and most are from the City of Sao Paulo itself.

2.2 Sampling

Three main sampling techniques were used: systematic sampling in the treatment services and a combination of snowball sampling and privileged access interviewer technique for the non-treatment sample. At the treatment sites an attempt was made to interview all available patients, but due to time constraints this was not always possible. In this situation patients were interviewed according to the order in which their names appeared on the inpatient lists or in order of arrival at the outpatient clinics. Data were not collected on patients who were not interviewed. A snowball sample of 68 cocaine users was obtained by asking users in treatment settings to nominate a cocaine using friend or colleague who was not currently in treatment. A further 27 out-of-treatment cocaine users were contacted mainly through non-drug users who knew people who

used cocaine within their own social or professional network, including a dentist, a porter and several drugs workers.

2.3 Procedures

In *Study I* 294 patients were interviewed by a team of 3 researchers using a questionnaire that had been developed and piloted in Brazil. In *Study II* 332 patients were interviewed by a team of 10 researchers using a questionnaire that had been translated from English. In both studies patients gave verbal consent and were guaranteed confidentiality as well as anonymity due to the very private and often illegal nature of the activities they were being asked about. Both projects were approved by the Ethics Committee.

Of the 23 sites included in the study three were visited by both research teams, although rarely at the same time. Checks were made to see if duplicate cases had been included by comparing the dates of birth of patients interviewed at these sites. Patients with identical birthdays had their questionnaires compared to see if their main demographic details were also identical (sex, level of schooling, marital status and profession). Three duplicate cases were found and excluded.

Compared to *Study I*, *Study II* had the stricter entry criteria: use of cocaine at least twice per week for three months vs. lifetime use of cocaine on more than one occasion. When the stricter criteria were applied to the *Study I* sample, 9 patients had to be excluded as they had never used cocaine with such a high frequency. The final sample size after exclusions was 614.

2.4 Statistics

For statistical purposes sites that shared common characteristics were grouped together into the following categories: public teaching hospital outpatient clinics (n=3), other public outpatient/counselling services (n=5), public hospital inpatient units (n=2),

HIV treatment services and hospices (n=8), private inpatient units (n=5) and an out-of-treatment sample. The two continuous variables that were investigated (age and duration of cocaine use) were skewed to the right and underwent transformation. A logarithmic transformation was used for age (transformed mean=3.25, S.D.=0.28, range 2.37 to 4.04) and a square-root transformation for duration of cocaine use (transformed mean=2.59, S.D.=0.99, range 0.40 to 6.16). All comparisons are between the six grouped sites. Age and duration of cocaine use were analysed using one-way analysis of variance. Gender differences were investigated using the X^2 test. Because there were substantial differences between sites in relation to age and duration of cocaine use, all other binary variables were analysed using logistic regression analysis so that the effects of these variables could be controlled for. In general, demographic variables were controlled for the effects of age, whilst drug and crime related variables were controlled for the effects of duration of cocaine use. Since all the statistical tests undertaken in this study were conceptually linked, a Bonferroni correction was made. Sixteen statistical tests were performed and the adjusted significance level was $p=0.0031$. Some variables were not tested because the vast majority of subjects shared that characteristic. This was true for lifetime use of the following substances: cannabis (96%), snorted cocaine (96%), alcohol (91%) and cigarettes (86%). As frequent cocaine use was used as a criterion for entry into the study, this variable was not submitted to statistical analysis.

3. Results

Six hundred and fourteen cocaine users were eligible for inclusion in the analysis. The differences between users in relation to the site at which they were interviewed are shown in Table 2. In terms of demographic characteristics, age and sex are the most strikingly different between sites. Cocaine users in the HIV treatment services were on average 8 years older than those in the public outpatient clinics.

Previous episodes of drug misuse treatment were very infrequent in the HIV services and not surprisingly in the out-of-treatment sample, whilst two-thirds of those interviewed in inpatient units had had prior treatment. Prostitution, or rather exchanging sex for drugs or money to buy drugs, showed a six-fold variation between sites being highest among those in HIV treatment centres. Criminal activity was also highest among patients seen in HIV services as well as those in public inpatient units. However, drug dealing was equally common among cocaine users from each principal site. In terms of cocaine history, patients seen in HIV services had the highest frequency of lifetime injecting of cocaine. Duration of cocaine use varied substantially between sites with those in HIV services having used for twice as long as those in public outpatient units.

4. Discussion

This study shows that cocaine users interviewed at diverse sites had markedly different characteristics, in terms of demographic features, criminal history, history of prostitution, previous drug treatment and drug use history. This suggests that by increasing the number of sites used for patient recruitment, and by including non-treatment settings, we can obtain more heterogeneous samples of drug users.

In a study similar to ours, Lewis et al (1992) interviewed 1,003 crack abusers from six types of locale in New York City: streets where drug users do their business, jails, probation/parole services, prisons, drug abuse treatment settings and courts. In the analysis the authors focus on differences in demographic features, including sex, age, race, education, marital status and employment, but not on patterns of drug use. The sampling strategy used in their study differed from ours in that predetermined quota targets were used to make sure that certain subgroups were not under sampled, in particular females and adolescents. Consequently, it is difficult to compare results.

However, variation between locales was found, especially in relation to race and employment status. In the United Kingdom National Treatment Outcome Research Study (Gossop et al, 1998), 1075 drug users undergoing treatment at 54 agencies were compared in terms of treatment modality: inpatient, residential rehabilitation, methadone maintenance and methadone reduction. Substantial differences were found between groups, in particular in relation to age, duration of heroin use, criminal history, treatment history and drug use history. As this was a treatment outcome study, there were no out-of-treatment subjects.

Although we used multisite sampling in our study, there was an over representation of cocaine users from treatment sites, in particular public treatment agencies. One might predict that this would tend to reduce heterogeneity and amplify the differences between treatment and non-treatment samples. However, this does not appear to have been the case, indeed the greatest absolute differences between variables occurred between treatment agencies. Of the 16 variables studied there will be a total of 32 upper and lower limits to these values. Eighteen of the most extreme values came just two sites (HIV services and public outpatient clinics), but each of the six sites provided at least three extreme values, suggesting that all were contributing to the increased heterogeneity.

Some potential sites where large numbers of drug users might have been found, such as prisons and illicit drug markets (Lewis et al, 1992), were not included in this study. A decision was taken during the planning stage not to interview drug users at these sites, due to a lack of safety.

Although multisite sampling appears to increase patient heterogeneity, we cannot be sure that it makes samples more representative. The relationship between heterogeneity and representativeness is unclear. Because selection bias operates in

different ways at different sites, we may have inadvertently over or under sampled certain subgroups of cocaine users, such as men and younger cocaine users. This could have affected the overall representativeness of the sample. The only way we could have avoided this is if we had used quota sampling developed from a knowledge of the characteristics of the sampling frame, i.e. all cocaine users in São Paulo. However, this information was and still is not available.

This analysis shows that multisite sampling is a valid procedure for increasing patient heterogeneity, but without information on the sampling frame it cannot be shown to improve representativeness. Since representativeness is the key to generalisability, this is a serious limitation. However, in different situations heterogeneity or representativeness may be more or less important. For example, in a study investigating the relationship between route of administration of a drug and severity of dependence (Gossop et al, 1992), the more heterogeneous the sample in relation to these variables the better. However, in a study measuring the prevalence of HIV among drug users (Haw et al, 1992), greater representativeness would take precedence. The relative importance of heterogeneity or representativeness will depend on the hypothesis being tested and the study design.

Acknowledgement

Dr J Dunn received a fellowship from the Brazilian Research Council (CNPq) and his work was in part funded by a grant from FAPESP. Dr C. P. Ferri received a fellowship from CAPES and part of her work was also funded by FAPESP. We gratefully acknowledge the statistical advice from Bob Blizard, statistician, Academic Dept. of Psychiatry, Royal Free Hospital.

References

- Carrol, K.M. , Rounsaville, B.J., 1992. Contrast of treatment-seeking and untreated cocaine abusers. *Arch. Gen. Psychiatry* 49, 464-471.
- Deren, S., Beardsley, M., Coyle, S., Singer, M., 1998. HIV serostatus and risk behaviors in a multisite sample of drug users. *J. Psychoactive Drugs* 30, 239-245.
- Dunn, J., Ferri, C.P., 1999. Epidemiological methods for research with drug misusers: review of methods for studying prevalence and morbidity. *Rev. Saude Publica* 33, 206-215.
- Dunn, J., Laranjeira, R.R., Silveira, D.X., Formigoni, M.L.O., Ferri, C.P., 1996. Crack cocaine: an increase in use among patients attending clinics in Sao Paulo 1990-1993. *Subst. Use Misuse* 31, 519-527.
- Dunn, J., Laranjeira, R., 1999a. Cocaine - profiles, drug histories, and patterns of use of patients from Brazil. *Subst. Use Misuse* 34, 1527-1548.
- Dunn, J., Laranjeira, R., 1999b. Transitions in the route of cocaine administration - characteristics, directions and associated variables. *Addiction* 94, 813-824.
- Edlin, B.R., Irwin, K.L., Faruque, S., McCoy, C.B., Word, C., Serrano, Y., Inciardi, J.A., Bowser, B.P., Schilling, R.F., Holmberg, S.D. & The Multicenter Crack Cocaine and HIV Infection Study Team, 1994. Intersecting epidemics - crack cocaine use and HIV infection among inner-city young adults. *N. Engl. J. Med.* 331, 1422-1427.
- Ferri, C.P., Gossop, M., 1999. Route of cocaine administration; patterns of use and problems among a Brazilian sample. *Addict. Behav.* 24, 815-821.
- Farrell, M., Howes, S., Taylor, C., Lewis, G., Jenkins, R., Bebbington, P., Jarvis, M., Brugha, T., Gill, B., Meltzer, H., 1998. Substance misuse and psychiatric

- comorbidity: an overview of the OPCS National Psychiatric Morbidity Survey. *Addict. Behav.* 23, 909-918.
- Gossop, M., Griffiths, P., Powis, B., Strang, J., 1992. Severity of dependence and route of administration of heroin, cocaine and amphetamines. *Br. J. Addiction* 87, 1527-1536.
- Gossop, M., Marsden, J., Stewart, D., Lehmann, P., Edwards, C., Wilson, A., Segar, G., 1998. Substance use, health and social problems of service users at 54 drug treatment agencies. Intake data from the National Treatment Outcome Research Study. *Br. J. Psychiatry* 173, 166-171.
- Griffin, M.L., Weiss, R.D., Mirin, S.M., Lange, U., 1989. A comparison of male and female cocaine abusers. *Arch. Gen. Psychiatry* 46, 122-126.
- Griffiths, P., Gossop, M., Powis, B., Strang, J., 1993. Reaching hidden populations of drug users by privileged access interviewers: methodological and practical issues. *Addiction* 88, 1617-1626.
- Haw, S., Frischer, M., Green, S., Crosier, A., Hunter, G., Finlay, A., Covell, R., Ettore, B., Bloor, M., Stephens, S., Goldberg, D., Stimson, G., McKeganey, N., Platt, S., Taylor, A., Follet, E., Parry, J., 1992. The importance of multistage sampling in determining the prevalence of HIV among drug injectors in Glasgow and London. *AIDS* 6, 517-518.
- Kleinman, P.H., Miller, A.B., Millman, R.B., Woody, G.E., Todd, T., Kemp, J., Lipton, D.S., 1990. Psychopathology among cocaine abusers entering treatment. *J. Nerv. Ment. Dis.* 178, 442-447.
- Lewis, C., Johnson, B.D., Golub, A., Dunlap, E., 1992. Studying crack abusers: strategies for recruiting the right tail of an ill-defined population. *J. Psychoactive Drugs* 24, 323-336.

Lopes, C.S., Lewis, G., Mann, A., 1996. Psychiatric and alcohol disorders as risk factors for drug abuse. *Soc. Psychiatry Psychiatr. Epidemiol.* 31, 355-363

Rounsaville, B.J., Kleber, H.D., 1985. Untreated opiate addicts. *Arch. Gen. Psychiatry* 42, 1072-1077.

Table 1. Characteristics of the 24 individual sites that were used in this study

Site name	Type of Service	Status	Clientele	Number Interviewed
UNIAD	Outpatients	Public	Drugs & alcohol	131
PROAD		Teaching		17
Santa Casa		Hospitals		1
PROSAM	Outpatient	NGO*	Drugs	30
DENARC	Counselling	Police	Drugs	40
NAT	Outpatients	Public	Drugs & alcohol	6
PAM		Municipal		14
CRTF		Clinics		15
Agua Funda	Inpatient units	Public	Drugs & alcohol	64
Taipas		Hospitals		73
CRAIDS	Outpatient	Public	HIV/AIDS	33
DIPA	HIV services	Clinics		9
Taipas-HIV				1
Brenda Lee	Hospices & nursing homes for people with HIV	NGOs or Charitable	HIV positive drug users, prostitutes, transvestites and transsexuals	6
Lar Betânia				6
Associação Paz				2
Projeto Esperança				2
Associação Lar				10
Instituto Bairral	Inpatient units	Private	Drugs & alcohol	15
Greenwood				7
Vila Serena				8
Nossa Senhora			General psychiatric	11
Joao de Deus				19
Non-treatment	Snowball	-	Regular cocaine users	68
	PAI†			27

*Non-governmental organisation

†Privileged access interviewer technique.

Table 2. Differences between cocaine users interviewed at six grouped sites in relation to demographic characteristics and drug use history (n=614).

Variable	University Outpatient (n=149) %	Public Outpatient (n=105) %	Public Inpatient (n=137) %	HIV Service (n=68) %	Private Inpatient (n=60) %	Non- treatment (n=95) %	p value††
Age - mean (S.D.) -yr. †	26.4 (7.57)	23.5 (7.29)	26.4 (6.96)	31.4 (6.42)	29.1 (9.32)	26.8 (7.58)	p<0.0001*
Sex - male	96	93	92	79	76	88	p<0.0001*
Marital status: single	65	76	57	46	58	75	p=0.058
married/other	36	24	43	54	42	27	
Schooling: 1 ⁰	61	80	70	69	58	77	p=0.019
2 ⁰ or above	39	20	30	31	41	23	
Employed	40	34	34	19	43	30	p=0.003*
Crime: Theft	38	52	61	60	45	58	p=0.0006*
Armed robbery	15	18	32	32	10	32	p<0.0001*
Drug dealing	30	29	33	37	32	40	p=0.475
Arrested	48	39	56	85	46	48	p<0.0001*
Prostitution	4	6	14	25	20	24	p<0.0001*
Past treatment	46	42	65	24	63	20	p<0.0001*
Lifetime use of: Amphetamine	19	10	19	38	30	11	p=0.009
Tranquillisers	38	35	51	56	38	24	p=0.0007*
I.V. cocaine	17	13	20	78	28	27	p<0.0001*
Crack	85	83	92	76	75	82	p=0.053
Duration of cocaine use (SD) -yr. †	7.1 (5.18)	6.1 (5.20)	7.5 (5.60)	11.0 (5.50)	9.5 (7.80)	7.7 (4.79)	p<0.0001*

* with Bonferroni adjustment, significance level is p=0.0031

† Age underwent a natural logarithmic transformation and *duration of cocaine use* a square-root transformation prior to analysis.

†† All analyses were corrected for the effects of age or duration of cocaine use, except age and *duration* themselves and *sex*.