The Analysis of Time-location Sampling Study Data*

John M. Karon
Emergint Corporation, Louisville, KY 40206

Abstract

Time-location sampling (TLS) is used to collect information from hard-to-reach populations by sampling persons at locations at which they may be found. Epidemiologic studies using TLS have often been analyzed ignoring both clustering within locations and the differential probabilities that persons are sampled. I propose a weighted analysis reflecting approximate differential sampling probabilities that may permit generalization to the population of persons attending locations in the sampling frame. I illustrate the effects of clustering and weighting using TLS data from a study of men who have sex with men. Some design effects are large. In designing a TLS, investigators should think carefully about collecting information about attendance at sampling locations, so that it may be possible to estimate a participant's probability of being sampled.

Key words: time-location sampling; time-space sampling; multiplicity problem; design effect

1. Introduction

Time-location sampling (TLS; also known as time-space sampling) has been used to collect data from men who have sex with men (MSM) by sampling at locations where these men can be found. Published analyses of these data have regarded the data as if they were obtained from a standard epidemiological cross-sectional study. Such analyses ignore the facts that the probability of being invited to participate varies among men and that characteristics of interest may be correlated within sampling locations. Properly accounting for both of these sample characteristics affects variance estimates. Properly accounting for sampling probabilities may also affect point estimates.

It is very difficult to estimate a person's sampling probability for a TLS study of MSM. I suggest an estimate, based on data which it is feasible to collect, that is approximately proportional to this sampling probability under certain assumptions. The inverse of this estimate can be used as a weight for a weighted analysis. I use data from a TLS of MSM conducted by the Centers for Disease Control and Prevention (CDC) to demonstrate the variance inflation that may occur by taking into account weighting and clustering (within sampling locations). Some point estimates in this study are sensitive to whether the analysis is weighted. Detailed exploratory analyses can help explain the differences between weighted and unweighted analyses, but caution must be used in interpreting an estimate for which the two analyses give differing results.

As discussed below, it is probably impossible to estimate the probability that an MSM will be sampled in a time-location study. Therefore, it is not possible to estimate the size of the population from which the sample is drawn. My comments are restricted to estimating a mean.

2. Construction of a Time-location Sample

Time-location sampling is used to sample a population for which a sampling frame cannot be constructed but locations are known at which the population of interest can be found, or for which it is more efficient to sample at these locations. Such populations include homeless persons, migrant workers, museum attendees, and blood donors. Kalton (1991) gives examples of TLS studies and practical advice on conducting such a study. MacKellar et al. (1996) give information on the use of TLS for studies of MSM.

I assume that sampling is to be done at more than one location. The sampling frame is the locations at which there is sufficient attendance by persons in the population of interest to make sampling worthwhile. A random sample of locations is chosen from this frame. If attendance depends on day of the week and time of day (as it does in studies of MSM), a sampling period is then chosen for each location in the sample. If locations vary in the frequency with which they have the necessary number of attendees, some care is required in constructing the sampling calendar, such as first choosing sampling periods for locations with the fewest available periods. Finally, a sample of attendees is chosen during each sampling event. If possible, this should be a random sample. The investigators should estimate the sampling

fraction by recording the total number of persons at the location during the sampling period who meet, or appear to meet, the eligibility criteria for the study.

3. The Design Effect

I use Kish's definition of the design effect as the ratio of the variance of the estimate (based on both the estimator used and the sampling design) to the variance of the estimate based on a simple random sample of the same size selected with replacement (srswr) (Kish, 1992). For a weighted analysis of a ratio mean (e.g. a proportion) in a cluster sample, using the inverse of the sampling probabilities as weights, an approximate expression for this design effect is

$$Deft^2 = \{(1+(\overline{n}-1)\rho)(1+CV_w^2)\} \qquad (1)$$

where $\overline{n}$ is the mean size of the primary sampling units, ρ is the intracluster correlation coefficient (ICC) of the characteristic being estimated (such as a proportion), and the final term is the square of the coefficient of variation of the weights (Kish, 1987). The ICC is zero if the variability of the characteristic within clusters is the same as the variability in the population; the ICC is positive if the variability of the characteristic within clusters is less than the variability in the population. Note that we can calculate the ICC by using a statistical software package to obtain an estimate of the variance or standard error from an unweighted clustered analysis, computing the mean cluster size, and computing the variance or standard error obtained from a srswr.

Park and Lee (2004) showed that the final term in the expression for the design effect must be modified if the sampling probabilities are correlated with the ratio being estimated. Their expression for the design effect also depends on the square of the coefficient of the weights.

4. Weights for a TLS

There are differential sampling probabilities in a TLS because persons in the population of interest vary in the frequency with which they attend locations in the sampling frame. This is an example of the multiplicity problem because a person could be selected for the sample from more than one of the primary sampling units. In theory, we can compute the probability of inclusion in the study. Let p be the probability that a specific person is enrolled in the study. Let s_i be the probability that s/he is enrolled during the i^{th} sampling event, given his/her behavior (how often the person attends locations in the sampling frame, which location s/he attends, and when s/he attends). Then

$$p = s_1 + (1-s_1)s_2 + (1-s_1)(1-s_2)s_3 + \ldots$$

That is, the person could be enrolled during the first sampling event, or not enrolled during that event but enrolled during the second, or not enrolled during the first two events but enrolled during the third, etc.

CDC investigators attempted to estimate p in analyzing the Young Men's Study Phase I (Valleroy et al., 2000). As part of the interview, they obtained information on how often a man went to some venue in the sampling frame and the probability distribution of the types of venues he attended. Note that

$$s_i = \gamma_i \alpha_i f_i$$

where γ_i is the probability that a man attends some venue in the sampling frame on the i^{th} day when sampling takes place, α_i is the probability that he attends the venue at which sampling takes place, and f_i is the sampling fraction at that sampling event (assuming that every eligible person is willing to participate). CDC estimated s_i for each man based on the proportions of different types of venues in the sampling frame under the assumption that attendance at all venues of a particular type was equally probable. However, the numerical estimates of the sampling probability p were approximately 1 for nearly all men; this implies that nearly every man in the population was sampled, which clearly was not true (sampling fractions were less than 0.25 even at the end of the study).

Therefore, we seek an alternative approach. Suppose that, for an individual, s_i is constant and equal to s. For example, this will be true if each of γ_i, α_i, and the sampling fractions are constant during the study period; the α_i are constant if the person's probability of attending a venue is equal to the probability that the venue is sampled. Let n be the number of sampling events. Since the probability of not being sampled on each event is 1 – s, the probability of being sampled is

$$p = 1 - (1-s)^n = ns - n(n-1)s^2/2 + o(n^2s^2)$$

using the power expansion of $(1 + x)^n$. Assume that the population is large enough and the number of sampling events is small enough that ns is small. Then p is approximately proportional to γ. Thus, if s_i is approximately constant, equation (1) shows that we obtain approximately correct estimates of a mean (or proportion) by using the weight $1/\gamma$ for each person. The information necessary to estimate γ can be obtained in the interview. Sudman and Kalton (1986) proposed such a weighting procedure for sampling at a single location.

5. Analysis Example

CDC used time-location sampling to obtain information about young MSM in the Young Men's Study Phase II, a study conducted in six metropolitan areas during 1998 – 2000 of 2942 MSM ages 23 to 29 years. Men were recruited at nine types of venues, including bars, dance clubs, businesses, health clubs, sex establishments, social organizations, street corners, and parks. Investigators asked men how often they attended bars or dance clubs during the last 6 months: never, less than once per month, once per month, 2 to 3 times per month, once per week, 2 to 3 times per week, or daily. See MacKellar et al. (1996) for more information on the design of this study.

For illustration, I restrict analyses to four of the six cities. Table 1 shows the proportion of men with each frequency of attendance and the corresponding weights. The proportion of men recruited from bars and dance clubs varied greatly from approximately 40% to 85% among these metropolitan areas. However, the distribution of weights was very uniform across these four cities.

Table 2 shows estimates of HIV prevalence and standard errors for these four cities from alternative possible analyses. The naïve analysis is a standard epidemiologic analysis (unweighted and ignoring clustering), with the usual estimate of the standard error obtained from the binomial distribution. The standard errors for the clustered analyses are from Proc SurveyMeans in SAS version 9.1. The greatest effect of weighting on the point estimate is for city C. For each city except for city A, both clustered analyses have design effects of at least 2, and all but one of these is at least 3. Some design effects for weighted clustered analyses are less than those for unweighted clustered analyses.

Table 3 shows corresponding results for the prevalence of Hepatitis B, which is correlated with HIV, and for unprotected anal intercourse during the last 6 months, which is a risk factor for both HIV and Hepatitis B infection. Except for city A, the design effects for HIV and Hepatitis B are similar. For these analyses, the standard error for each weighted clustered analysis is greater than that for the unweighted clustered analysis.

Table 1. Analysis weight and percent of participants, by reported frequency of attendance at bars and night clubs, for four metropolitan areas in the Young Men's Study Phase II.

Frequency of Attendance	Weight	Percent of participants
Never	Deleted from analysis	2
< once per month	1 / .01 = 100	11
Once per month	1 / .03 = 33.3	10
2-3 times per month	1 / .08 = 12.5	25
Once per week	1 / .14 = 7.1	20
2-3 times per week	1 / .35 = 2.9	26
Daily	1 / 1 = 1	5

Table 2. Alternative analyses of HIV prevalence, and standard errors of the estimates, by city, for four cities in the Young Men's Study Phase II.

			clustered analyses	
City	Design effect	naive analysis	unweighted	attendance weights
A	4.8	.114±.015	.114±.017	.144±.033
B	4.6	.165±.016	.165±.027	.146±.039
C	4.7	.145±.016	.145±.048	.184±.035
D	5.0	.190±.018	.190+.024	.208±.040

The naïve analysis ignores clustering and weighting. The design effects are for the clustered weighted analyses.

Table 3. Alternative clustered analyses of the prevalence of Hepatitis B and unprotected anal intercourse, and standard errors for these estimates, by city, for four cities in the Young Men's Study Phase II.

	Hepatitis B			Unprotected anal intercourse		
city	design effect	unweighted	attendance weights	design effect	unweighted	attendance weights
A	2.6	.156±.019	.166±.028	4.2	.469±.022	.353±.032
B	3.3	.294±.026	.262±.036	3.1	.417±.021	.408±.040
C	6.5	.179±.035	.219±.044	3.1	.421±.021	.427±.042
D	4.7	.238±.019	.262±.043	1.2	.522±.014	.500±.033

The design effects are for the weighted clustered analyses.

Table 4. Design effects and factors affecting them, by city, for the four cities in the Young Men's Study Phase II analysis.

City	Design effect		Mean cluster size	ICC (p-value)		
	HIV	HBV		HIV	HBV	CV_w^2
A	4.8	2.6	11.2	.02 (.61)	.02 (.71)	2.1
B	4.6	3.3	13.3	.14 (.06)	.05 (.22)	2.3
C	4.7	6.5	25.4	.33 (.06)	.12 (.14)	2.1
D	5.0	4.7	18.5	.05 (.20)	-.00 (NC)	2.1

HBV: Hepatitis B virus; CV_w^2: square of the coefficient of variation of the analysis weights. NC: algorithm did not converge.
The p-values are from a logistic regression mixed model.

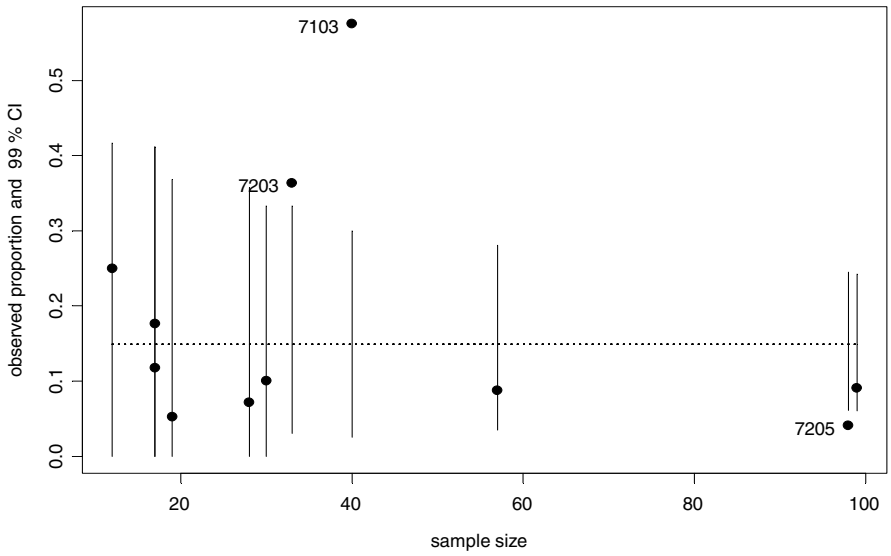
To understand the possible reasons for the great variation in design effects, I computed approximate values for the intracluster correlation coefficient (ICC) by using the relation in (1) corresponding to a clustered unweighted analysis. Table 4 shows the mean cluster sizes and estimates of the ICC for HIV prevalence in these cities. There is little variation in the CV of the weights among the cities. Both the mean cluster size and the ICC are substantially larger in city C than in the other cities.

It would be of interest to do a formal test for intracluster correlation. Table 4 also shows the p-values obtained from this test by fitting a logistic regression mixed model using Proc NLMixed in SAS version 9.1. This procedure allows only Gaussian variation among the prevalences at the venues. Even the p-value for HIV in City C is only marginally significant.

As an alternative to a formal test procedure, an exploratory graphic plot is much more useful. The Figure shows the observed HIV prevalence in city C at each venue from which at least 10 men were recruited. It also

shows a 99% binomial confidence interval for the prevalence in each venue under the assumption of no clustering and conditional on the venue sample size. The prevalences at venues 7103 and 7203 are very high; the prevalence at 7205 is lower than expected. HIV prevalence in African Americans in city C was much higher than in men of other races; nearly all the men enrolled at venues 7103 and 7203 were African American. At venue 7205, 11% of the men enrolled were African American, compared to 30% overall. This plot suggests that there is intracluster correlation, likely as a result of clustering by race/ethnicity.

Figure. 99% confidence intervals (vertical lines) for HIV prevalence by venue in city C, conditional on the number of men sampled at the venue, assuming the true prevalence in each venue was equal to the overall prevalence (14.9%, dashed line); dots are observed prevalences.



6. Logistic Regression Models for Associations

Clustering and variation in frequency of attendance at venues may also affect estimates from models. Table 5 shows the results from evaluating unprotected anal intercourse during the last 6 months (UAI) as a risk factor for HIV and for Hepatitis B in city C using logistic regression models. The clustered model results are from Proc SurveyLogistic in SAS version 9.1. For HIV as the outcome, weighting inflates the variance, but all point estimates are similar. For Hepatitis B as the outcome, the point estimate of the log odds ratio is very different in the clustered weighted model than in the unweighted models.

The reason for this dependence requires an analysis of the association between UAI and Hepatitis B within groups defined by reported frequency of attendance at bars and dance clubs. The Breslow-Day test for the uniformity of the odds ratio among these attendance groups is marginally significant (p = .06). The Mantel-Haenszel estimate of the odds ratio is 0.91, similar to the unweighted estimates. Table 6 shows the number of men and observed odds ratio in each of these groups. The effective total sample size (using the weights) is approximately 10,100; the men who reported going to bars and clubs at most once per month (24% of the actual sample) contribute more than 75% of the weighted sample size. Therefore, the weighted odds ratio estimate is close to the odds ratio for these men who went to bars and dance clubs infrequently.

Table 5. Estimates of unprotected anal intercourse during the last 6 months as a risk factor for HIV and Hepatitis B in City C from alternative logistic regression models.

Model	Odds ratio (95% confidence interval)	
	HIV	Hepatitis B
Naïve	1.19 (0.71 – 1.98)	0.88 (0.55 – 1.42)
Clustered, unweighted	1.19 (0.82 – 1.72)	0.88 (0.54 – 1.45)
Clustered, weighted	1.41 (0.22 – 3.60)	0.52 (0.22 – 1.22)

Table 6. Number of men and odds ratios for the association between Hepatitis B prevalence and unprotected anal intercourse during the last 6 months, by frequency of attendance at bars and clubs.

Frequency at bars, clubs	< once / Month	once / month	2-3 / month	Once / Week	2-3 / week	Every day
N	59	54	121	89	128	29
weight	100	33	12	7	3	1
Effective N	5900	1800	1450	525	375	25
Odds ratio	0.46	0.46	1.5	0.19	2.2	1.1

Effective sample sizes (number in the sample, times the attendance weight) are rounded.

7. Discussion

Time-location sampling is a convenient method for sampling some hard-to-reach populations. However, if the analysis is unweighted, we can only claim that estimates refer to the actual persons sampled. If sampling fractions are used to compute weights, or if the analysis is unweighted but the sampling fractions are approximately constant, estimates refer to the population of visits to the locations in the sampling frame. It is necessary to use weights based on the probability that a person is sampled, accounting for variation of the frequency with which persons attend a location in the sampling frame, to obtain an estimate that refers to the population of persons who attend these locations. In sampling a population such as MSM, good estimates of these sampling probabilities are likely to be hard to obtain. For either a weighted or an unweighted analysis, the analysis should include the effect on standard errors of clustering within sampling locations (and perhaps also within sampling events at each location).

As a result of these considerations, investigators should collect information on the frequency with which persons in the population of interest attend venues in the sampling frame. Because good information is difficult to collect, questions concerning attendance should be phrased carefully. For example, it may be useful to ask how frequently a person attends each type of venue in the sampling frame. An approximate frequency of attending a sampled venue might be computed as a weighted average of these venue-type frequencies, using as weights the proportion of each type of venue in the sampling frame. It would be less useful to ask only how often a person attends some venue in the frame.

The data analysis should include an evaluation of associations between frequency of attendance and both outcomes of interest and important covariates. If there are clear associations, appropriate care should be used in drawing conclusions from the study. Alternatively, if weights can be defined which reflect the probability of being sampled for the study, both weighted and unweighted analyses should be done. If there appears to be a meaningful difference between the results, care must be used in drawing conclusions. Of course, all analyses should take into account clustering within venues.

As the analyses of the Young Men's Study Phase II show, design effects may be large (even for an unweighted analysis) and probably cannot be estimated in advance. Sample sizes for a time-space sample must take into account the possibility of large design effects.

Other methods have been proposed for sampling hard-to-reach populations, particularly snowball sampling and, recently, respondent-driven sampling (Heckathorn, 1997). A disadvantage of TLS is that inference is limited to the population of persons attending venues in the sampling frame. TLS is also likely to be

more expensive than these alternative methods. In any of these methods, it is hard to estimate relative probabilities that persons are sampled. At this time, TLS does have two advantages over these alternative methods. Statistical methods for data analysis of a TLS study are well understood and have a firm theoretical foundation. In addition, study participants themselves select most of the participants in snowball sampling or respondent-driven sampling, which may lead to biased results. This source of bias is eliminated in TLS because investigators select participants.

References

Heckathorn DD (1997). Respondent-driven sampling: a new approach to the study of hidden populations. Social Problems, 44, 174 – 199.

Kalton G (1991). Sampling flows of mobile human populations. Survey Methodology, 17, 183-194.

Kish L (1987). Weighting in Deft. The Survey Statistician, June 1987.

Kish L (1992). Weighting for unequal p_i. Journal of Official Statistics, 8, 183 – 200.

Park I, Lee H (2004). Design effects for the weighted mean and total estimators under complex survey sampling. Survey Methodology, 30, 183 – 193.

MacKellar D, Valleroy L, Karon J, Lemp G, Janssen R (1996). The Young Men's Survey: methods of estimating HIV seroprevalence and risk factors among young men who have sex with men. Public Health Reports 111 (Supplement 1), 138-144.

Sudman S, Kalton G (1986). New developments in the sampling of special populations. Ann. Rev. Sociol., 12, 401-429.

Valleroy L, MacKellar DA, Karon JM, Rosen DH, et al. (2000). HIV prevalence and associated risks in young men who have sex with men. J. Amer. Med. Assoc., 284, 198 – 204.

**The findings and conclusions in this report are those of the author and do not necessarily represent the views of the Department of Health and Human Services or the Centers for Disease Control and Prevention.*