

Washington University School of Medicine

Digital Commons@Becker

Posters

2003: Drinking and the High School Student

2003

Transect sampling methods for a minority population genetic epidemiology study

Nathan K. Risk

Krista L. Russell

Rumi Kato Price

Follow this and additional works at: <https://digitalcommons.wustl.edu/guzeposter2003>



Part of the [Medicine and Health Sciences Commons](#)

Recommended Citation

Risk, Nathan K.; Russell, Krista L.; and Price, Rumi Kato, "Transect sampling methods for a minority population genetic epidemiology study" (2003). *Posters*. Paper 13 Samuel B. Guze Symposium on Alcoholism.

<https://digitalcommons.wustl.edu/guzeposter2003/13>

This Poster is brought to you for free and open access by the 2003: Drinking and the High School Student at Digital Commons@Becker. It has been accepted for inclusion in Posters by an authorized administrator of Digital Commons@Becker. For more information, please contact vanam@wustl.edu.



Transect Sampling Methods for a Minority Population Genetic Epidemiology Study

Nathan K. Risk, M.A.

Krista L. Russell, B.A.

Rumi Kato Price, Ph.D., M.P.E.

*This study is supported by the Missouri Alcohol Research Center (MARC).
(P50AA1198, Center Director Andrew C. Heath, D. Phil)

Abstract

The Saint Louis Asian American Pilot Study is conducted under the auspices of the Missouri Alcohol Research Center (MARC). It proposes to conduct a cross sectional assessment in two Asian sub-populations, the Japanese and the Vietnamese residing in the Saint Louis Area.

Current population-based sampling methods that require sampling frames are unlikely to be suitable for efficiently sampling members of a small population. The population distribution in the Saint Louis area is not large enough to make the census-based sampling method reasonable. Although there are a number of other sampling methods that do not require sampling frames, the “transect” sampling derived originally from wildlife biology is a particularly flexible method.

This presentation will show: 1) the need for ascertaining the Asian sub-population in the United States; 2) the basics of the transect sampling method in wild-life biology; 3) the adaptation of the transect sampling method for the Saint Louis Asian population; and 4) simulated data that examines the robustness of transect sampling under various assumptions.

Introduction (1)

In **Table 1**, the rates of current drinking are reported separately for people of Asian decent who self identify as white versus people of Asian descent who self-identify as Asians using NLAES¹. The results show that the rates of drinking are higher among Asians who self identify as white². In **Table 2**, the rates of current drinking are reported separately for unmixed and mixed Asian adolescents using Add Health.³ The results consistently show that the rates of current drinking are higher among Asians of mixed heritage². It is an aim of this pilot study to examine the relative strength of acculturation measures such as mixed heritage on substance use and problem use in these two community samples.

The deficiency of the ALDH2 isozyme is known to cause a high sensitivity to alcohol among Asians⁴. This sensitivity is known as the “flushing syndrome.” There is also some evidence for a correlation of the wild type CYP2A6*1 with tobacco dependence⁵. Mutations of CYP2A6 have been reported at a higher frequency among Asians^{6,7}. Genotypic frequencies are given in **Table 3**. It is an aim of the pilot study to assess the feasibility of genotype collection for these two alleles for the Japanese and Vietnamese populations in Saint Louis, since genotypic distributions among Asians in the U.S. are expected to be substantively different from those in their native lands given the high rate of inter-racial reproduction currently on-going in the U.S.

**Table 1. Drinking by Racial Identity (%)
(Asian n=922)**

Race	Japanese	Filipino	Chinese	Korean	Vietnamese
Asian	30.4	31.1	15.3	##	##
White	46.9*	30.0	39.8*	##	##

Source: NLAES, 1992, Price, et. al., 2002. *, rate significantly higher among those identified themselves as whites; ##, unreliable estimates due to small numbers of Asians who identified themselves as whites.

Table 2. Drinking among Adolescent Asians by Multiple Racial Identity¹ (Asian n = 4,012)

Racial identity:	Japanese	Filipino	Chinese	Korean	Vietnamese
Unmixed Asian	24.5	24.3	9.7	17.1	12.9
Mixed Asian	39.2*	40.6*	41.7*	33.7*	49.4*

Source: Add Health S-95, Price et al, 2002. 1. Mixed Asian are those who reported at least one more race. Multiple choices for Asian ethnicity allowed. Weighted to be generalizable to the U.S. Population of adolescents in grade 7 through 12 in 1994-6. Standard errors adjusted using SUDAAN. *, significantly larger than unmixed Asians.

Table 3. Genotype Distributions of ALDH2 and CYP2A6

	ALDH2			CYP2A6		
	Genotypes (%)			Allelic (%) Frequency		
	*1/*1	*2/*1	*2/*2	6*1 ¹	6*2(v1) ²	6*3(v3)
Caucasian	100 ³	0	0	85 ⁴	15 ⁴	0 ⁴
Japanese	56.4	39.4	4.2	98.6 ⁵	1.4 ⁵	?
Filipino	87.3	12.7	0	Not Available		
Korean	71.6	26.6	1.8	Not Available		
Chinese	59.0	35.9	5.1	83 ⁶	11 ⁶	6 ⁶
Vietnamese	43.0	57.0		Not Available		

Source Harada, 1991; Goedde et al., 1985; 1. Wildtype 2. Inactive 3. European
4. Finnish based on the PCR amplifiers confined with diagnostic restriction
digestion. 5. Finnish, based on a two-step PCR method 6. Taiwanese

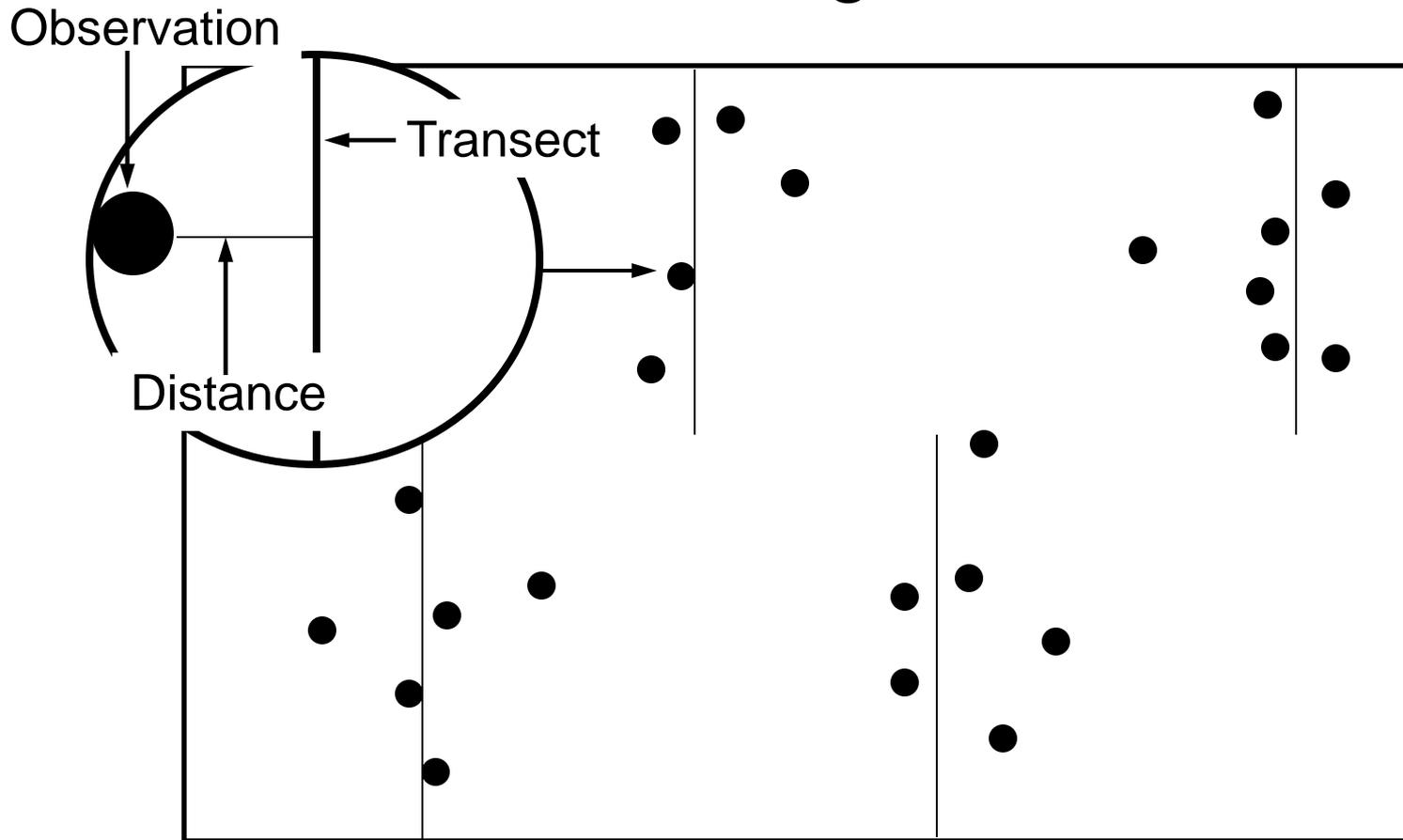
Introduction (2)

In any study of a minority population, ascertainment of a representative sample is a challenge. It is possible to ascertain a sample randomly by selecting blindly from the total population and accepting only respondents from the minority in question, but such an approach is prohibitively expensive. It is also tempting to enter the minority community at a few points and select respondents from a few well known centers of the community. Such an approach may well succeed in meeting the ascertainment goal, but the sample may not reflect the minority community under study as a whole.

In wild-life biology, transect sampling is used to ascertain a representative sample of the population of a given species by traveling along randomly chosen paths or “transects”. Observations of the species are noted and the distance to the observation from the transect is recorded (**Figure 1**).

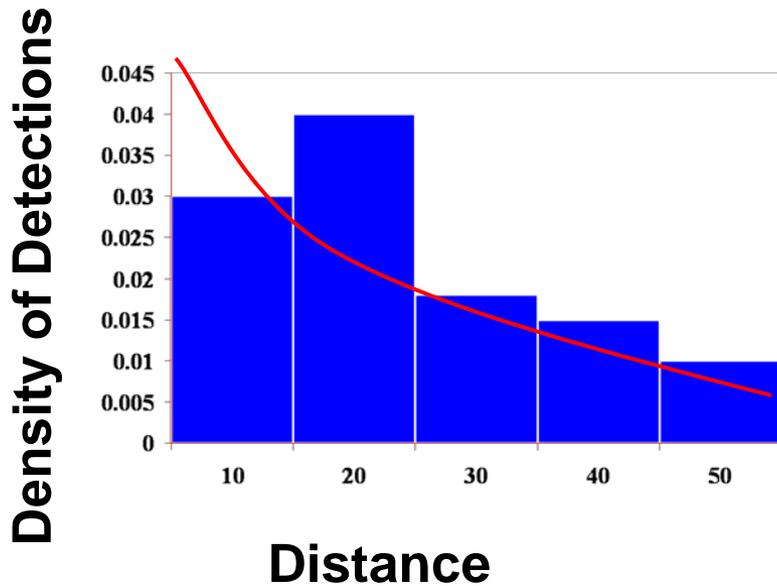
Since members of a species further from the transect are less likely to be spotted, the density of the species is estimated by fitting the percentage of observations made to distance. Several parametric methods for this fitting exist⁸. (**Figure 2**)

Figure 1. Transect Sampling Along Paths

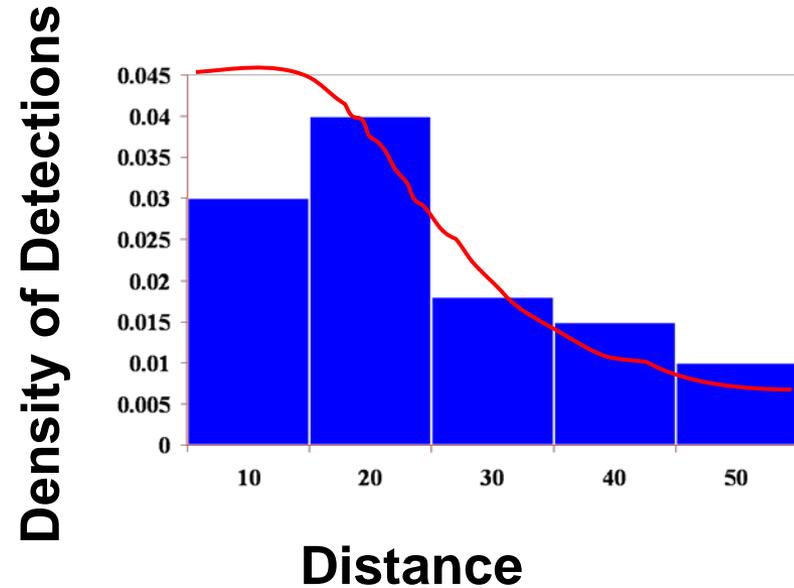


○ Blow-up detail of an observation and a transect.

Figure 2. Methods of Density Estimation in Transect Sampling



Exponential



Half - Normal

Source: Thompson, 1992.

Method (1)

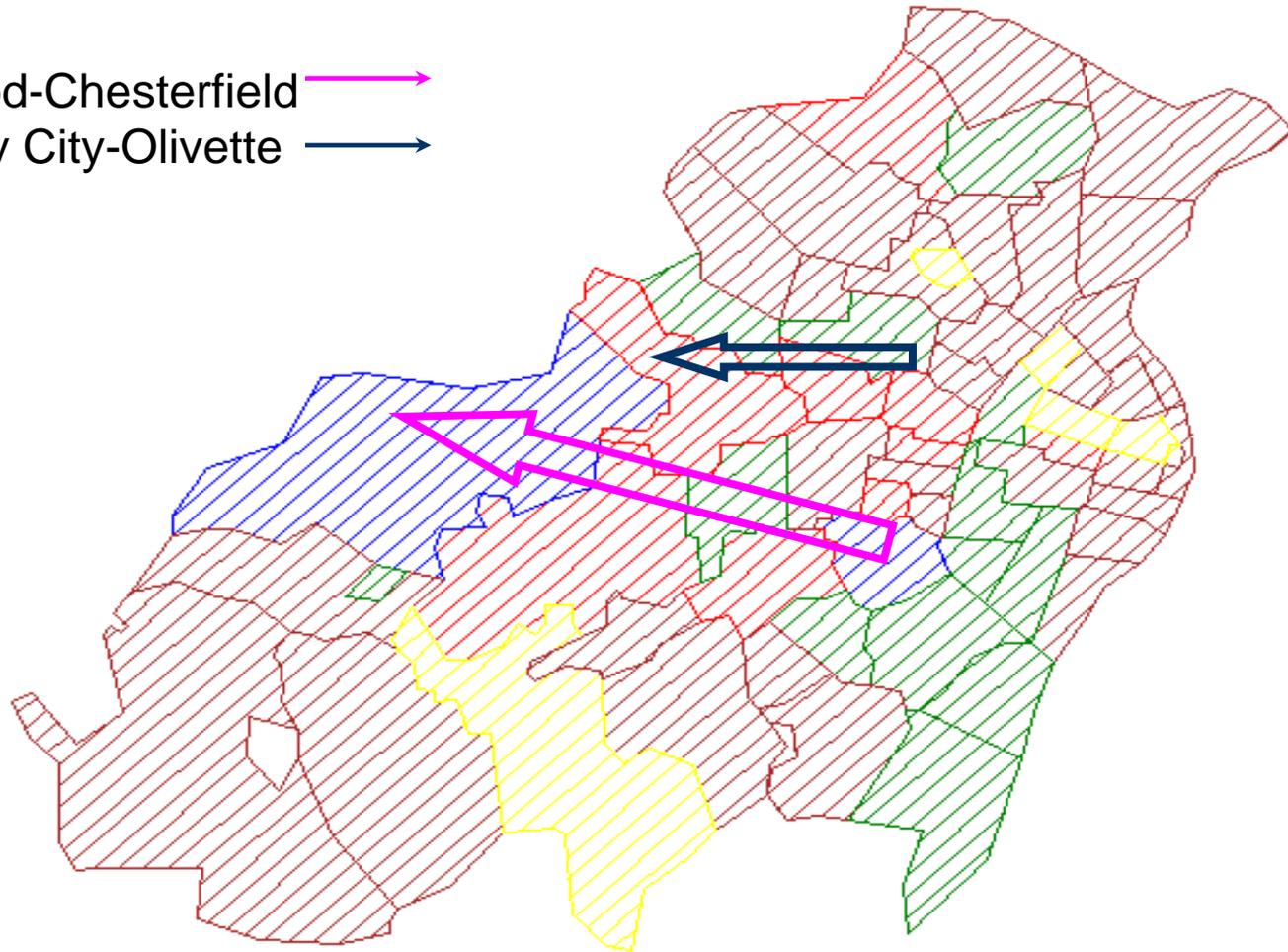
This pilot study adapts a transect sampling method for the purpose of ascertaining the Japanese and Vietnamese samples residing in Saint Louis (**Table 5** on sister poster.)

The Japanese sub-population is centered in two bands, Brentwood to Chesterfield and University City to Olivette (**Figure 3**). The Vietnamese population is also centered in two bands, Olivette to Maryland Heights and South City to South County (**Figure 4**).

The entry points into this population will be by community organizations and retail services that serve the minority community. Instead of the path method commonly used in transect sampling, each entry point into the community will recruit by advertisement. For the ascertained sample to successfully represent the community, the entry points must be scattered thoroughly throughout the community. By selecting a large variety of community organizations and retail services (**Figure 5**), the pilot study hopes to ascertain a representative sample from the respective populations in a cost-efficient manner.

Figure 3. Distribution of the Japanese Population in Saint Louis

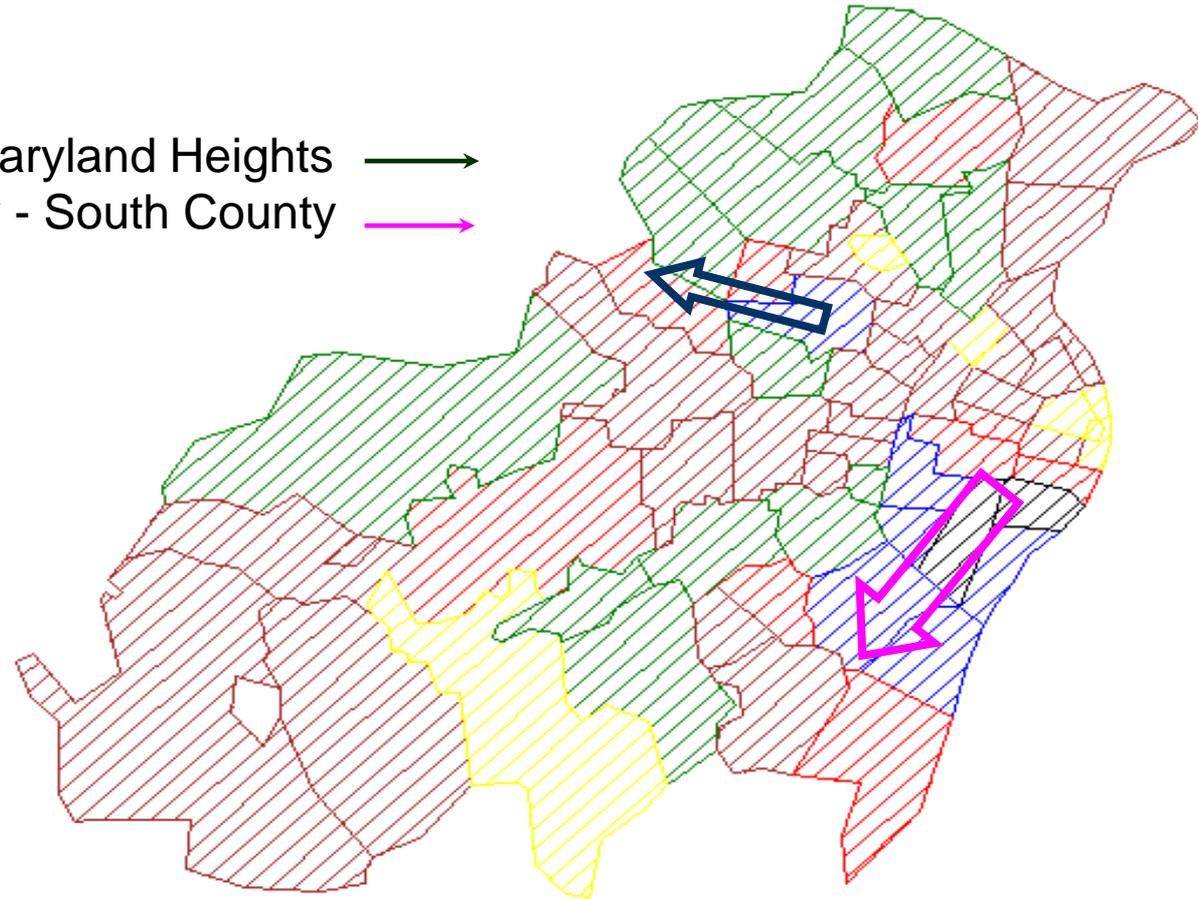
Brentwood-Chesterfield →
University City-Olivette →



JAPANESE 0 1 to 24 25 to 49 50 to 99 100+

Figure 4. Distribution of the Vietnamese Population in Saint Louis

Olivette-Maryland Heights →
South City - South County →



VIETNAM

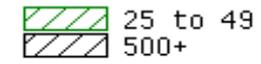
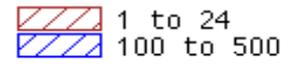


Figure 5. Map of Community Organizations and Service Providers¹



1. The list is recompiled periodically.

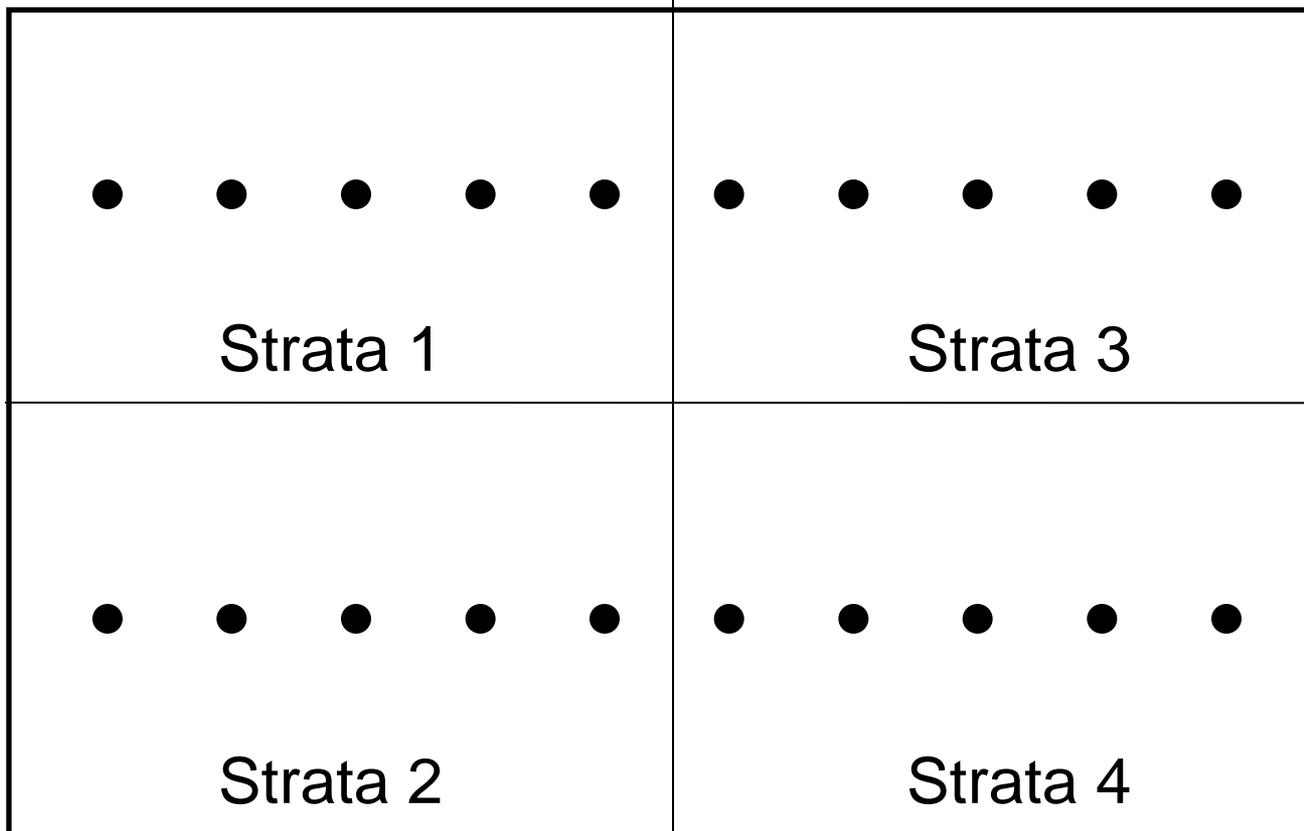
Method (2)

In order to test the effectiveness of transect sampling, several simulations were performed. The basic simulation was a four-strata simulation. The strata were geographically assigned to the four corners of the sample space (**Figure 6**). This mimics the actual geographic layout of the Japanese residing in Saint Louis County.

Twenty entry points were placed among the strata. Each entry point was given a randomly assigned popularity score. Observations were randomly assigned to a location inside of their strata and each observation was randomly assigned a binary score (such as mixed-heritage or high-school education) for Traits A and B based on the strata they were in.

For each entry point, a person (observation) would randomly visit that entry point based on the popularity of the entry point and the distance of the person from the entry point. A person visiting any entry point was considered to be sampled. For Simulation 1, all observations behaved similarly. For Simulations 2-5, observations from strata 1 were less likely to visit distant entry points and hence less likely to be sampled while observations from strata 2 were more likely to visit distant entry points and hence more likely to be sampled.

Figure 6. Four Strata Simulation with 20 Entry Points



Results

- For Simulation 1 (**Table 4**), when all members of the population respond similarly to the distance to entry points, each strata is sampled at a similar rate (18.6, 18.8, 18.6, 18.9) and therefore the sample proportion of Traits A and B is similar to the population proportion of Traits A and B (47.1 versus 47.3 and 32.1 versus 32.1 respectively)
- For Simulations 2-5, as strata 1 becomes sampled at an increasingly lower rate and strata 2 becomes sampled at an increasingly higher rate, the sample proportions of Traits A and B vary to an increasing extent from the population proportion of Traits A and B.
- The estimator of p_i , \hat{p}_i is generally adequate, but tends to be lower than p_i , since it is biased low by definition. An increased number of entry points and an even sampling across entry points will reduced this bias.
- Transect sampling is a reasonable method to obtain a representative sample if enough entry points are provided for each strata of the population. It is also feasible to measure the effectiveness of this sampling strategy for equality of sampling across strata.

Table 4. Simulation Results

	p_1	\hat{p}_1	p_2	\hat{p}_2	p_3	p_4	Trait A	Trait B
Population	—	—	—	—	—	—	47.3	32.1
Simulation 1	18.6	19.3	18.8	16.1	18.6	18.9	47.1	32.1
Simulation 2	16.9	16.8	20.9	18.4	18.6	18.9	47.0	33.9
Simulation 3	15.8	15.5	22.5	21.2	18.6	18.9	46.8	35.6
Simulation 4	14.9	14.3	24.2	23.5	18.6	18.9	46.4	37.5
Simulation 5	13.9	13.6	26.1	25.4	18.6	18.9	45.9	39.1

Traits A and B give the sampled proportion of Traits A and B in each simulation. p_i is the sampling proportion of strata i , (i.e. the proportion of observations that visited any of the 20 entry points.)

$$\hat{p}_i = \frac{2 \cdot (\text{number of observations sampled more than once in strata } i)}{(\text{number of observations sampled at least once in strata } i)}$$

\hat{p}_i is an estimate of p_i . Actually, \hat{p}_i is the probability of visiting any of 19 entry points instead of any of 20 entry points. Therefore it has a low bias for estimating p_i .

Literature Cited

1. National Institute on Alcohol and Abuse and Alcoholism. National Longitudinal Epidemiologic Alcohol Survey: Wave 1 Questionnaire. Rockville (MD): NIAAA; 1991.
2. Price, RK, et al. Substance Use and Abuse by Asian Americans and Pacific Islanders: Preliminary Results from Four National Epidemiologic Studies. Public Health Reports; 2002; 39-49
3. National Longitudinal Study of Adolescent Health. Research Design. Chapel Hill (NC): Carolina Population Center, Univ of North Carolina; 1998.
4. Crabb DW, et al. Genetic factors that reduce risk for developing alcoholism in animals and humans. In Begleiter H, Kissin B (eds.), The genetics of alcoholism. New York (NY): Oxford University Press, 1995; 202-220
5. Wellington C. Genes and tobacco dependence (a review). Clin Genet 1998; 54:266-267.
6. Yokoi T, Kamataki T. Genetic polymorphism of drug metabolizing enzymes: New mutations in CYP2D6 and CYP2A6 genes in Japanese. Pharmaceut Res 1998; 15:517-524.
7. Oscarson M, et al. Characterisation and PCR-based detection of a CYP2A6 gene deletion found at a high frequency in a Chinese population. FEBS Letters 1999; 448:105-110.
8. Thompson, S. Sampling. New York (NY): John Wiley and Sons, Inc., 1992