

Self-Collection of Oral Epithelial Cell DNA under Instruction from Epidemiologic Interviewers

Lea C. Harty,¹ Peter G. Shields,² Deborah M. Winn,³ Neil E. Caporaso,¹ and Richard B. Hayes⁴

Oral epithelial cells provide an easily accessible source of germline DNA. Two methods for collection were compared in a 1992–1995 case-control study of oral cancer in Puerto Rico. One group of subjects (55 controls without oral cancer) collected oral rinse samples at home or work under the direction of a nonmedically trained interviewer (“self-collection”); the other group (94 controls) participated in a clinic-based collection, which also included blood and urine samples, conducted by a medical technician (“clinic collection”). Participation was higher for self-collection (98.2%) than for clinic collection (70.7%) ($p < 0.001$). DNA yields ranged from 2.0 to 204.5 μg (median, 25.9 μg) and did not differ by collection method, although yields varied by interviewer among self-collected samples ($p = 0.02$). Success rates for polymerase chain reaction amplification of the *ADH3*, *NAT1*, and multiplex *CYP1A1/GSTT1/GSTM1* genotyping assays ranged from 76.4% (*NAT1*) to 98.2% (*ADH3*) for self-collected samples and were similar to those for clinic-collected samples (87.2–97.9%). Failure to amplify was associated with low DNA content ($p = 0.015$). Similar results were observed among cases (91 self-collected, 66 clinic collected), except that DNA yields did not vary by interviewer and a larger fraction (10.2%) of samples contained less than 5 μg of DNA, perhaps because of disease-related oral impairment. Self-collection of oral epithelial DNA samples appears satisfactory and efficient for many epidemiologic studies. *Am J Epidemiol* 2000;151:199–205.

DNA; epidemiologic methods; epidemiology, molecular; mouth mucosa; mouthwashes; polymerase chain reaction

Many epidemiologic studies investigate genetic factors as possible determinants of disease. Historically, DNA has been obtained from venous blood, but oral epithelial cells have recently been shown to be suitable for polymerase chain reaction (PCR)-based assays of subjects' genotypes (1–5). Oral epithelial cells are constantly exfoliated and may be captured through gentle scraping of the oral mucosa or by oral rinsing (1–5). Such collection methods are simple, are noninvasive, and avoid exposure to blood products. They are especially useful if subjects are very young; an inexpensive method for sampling large populations is needed;

blood collection is contraindicated for medical, cultural, religious, or personal reasons; or medically trained personnel and specially equipped facilities are unavailable.

Self-collection of oral epithelial cells at a subject's home or workplace under instruction from nonmedically trained interviewers is an attractive and efficient approach for many epidemiologic studies. We evaluated this approach and compared it with clinic-based collection by a medical technician on the basis of DNA quantity, suitability for PCR-based assays, and subject acceptability. We also investigated the influence of inter-interviewer variability and subject-related factors (age, gender, tobacco use, and alcohol drinking) on the collection results.

MATERIALS AND METHODS

Study design

In a population-based case-control study conducted in Puerto Rico from 1992 to 1995, we interviewed 367 patients with oral cancer and 521 age- and gender-matched controls with no history of oral cancer. Cases were ascertained through the Puerto Rico Central

Received for publication July 6, 1998, and accepted for publication April 14, 1999.

Abbreviation: PCR, polymerase chain reaction.

¹Genetic Epidemiology Branch, National Cancer Institute, Bethesda, MD.

²Laboratory of Human Carcinogenesis, National Cancer Institute, Bethesda, MD.

³Oral Health Promotion, Risk Factors, and Molecular Epidemiology Branch, National Institute of Dental Research, Bethesda, MD.

⁴Occupational Epidemiology Branch, National Cancer Institute, Bethesda, MD.

Reprint requests to Dr. Neil E. Caporaso, Genetic Epidemiology Branch, National Cancer Institute, Executive Plaza South Suite 7116, MSC 7236, Bethesda, MD 20892-7236.

Cancer Registry and local pathology laboratories. Controls were identified by using an area probability sampling frame (age <65 years) or by randomly sampling Medicare rosters (age ≥65 years). The study design has been described more fully elsewhere (6, 7).

A two-tiered approach was used to collect oral epithelial cells from interviewed subjects. For those from the San Juan area, a medical technician from the University of Puerto Rico Dental Clinic collected oral rinse, urine, and blood samples during a separate appointment within approximately 1 month of successful completion of the interview ("clinic collection"). During the 30-month field period, 98 cases and 133 controls were eligible for clinic collection. A total of 178 cases and 333 controls who resided outside the San Juan area and were interviewed during the first 19 months of the field period were not eligible to donate biospecimens.

To increase the sample size for genetic and viral studies during the final 11 months of the 30-month field period, a modified collection technique was introduced to obtain oral epithelial cells from subjects residing outside the San Juan area who had agreed to be interviewed. Ninety-one cases and 55 controls were eligible to collect oral rinse samples (but not blood or urine) themselves at their homes or workplaces under the direction of a nonmedically trained interviewer ("self-collection"). The interviewers were responsible for traveling to subjects' homes or workplaces throughout Puerto Rico to collect sociodemographic and potential risk factor information by using a questionnaire. During these 11 months, subjects (13 cases, 5 controls) residing in the San Juan area who refused to participate in the clinic collection were asked to perform self-collection, and 5 (4 cases and 1 control) did. When participation rates were calculated, these 18 subjects were included in the clinic collection group; when DNA yields and PCR results were analyzed, the five self-collected samples were included in the self-collected group.

All subjects gave written, informed consent to participate in the study, including oral epithelial cell donation. The study protocol was approved by the institutional review boards at the National Cancer Institute and the University of Puerto Rico. Subjects who participated in the clinic collection received up to \$40 for their travel-related expenses and inconvenience, whereas no monetary compensation was given for providing a self-collected sample.

Oral epithelial cell collection

Oral epithelial cell samples were collected by making at least 15 strokes on the oral mucosa and tongue with a soft-bristled cytobrush (Medical Packaging

Corporation, Camarillo, California), rinsing the mouth for 10 seconds with 10 ml of sterile water, and expectorating the rinse. For oral cancer patients, areas of tumor or prior surgery were avoided. In the clinic collection, the medical technician used gauze to hold the subject's cheeks and tongue while she brushed these tissues. Subjects participating in self-collection brushed their own cheeks and tongue, without stabilizing these structures, while the interviewer read step-by-step instructions and corrected deviations from the protocol. Both collection procedures took approximately 10 minutes to complete. After collection, the medical technician or interviewer mixed the oral rinse with 10 ml of 2× Standard Transport Medium (Digene Diagnostics, Inc., Silver Spring, Maryland). Collected volumes ranged from 5.5 to 24.5 ml, including the Standard Transport Medium. The medical technician and the interviewers received similar training on the proper technique for collecting oral epithelial cells, which consisted of attending a 1-day training workshop comprised of a demonstration, detailed written instructions, and hands-on practice in conducting satisfactory collections.

The clinic-collected samples were stored in a Styrofoam cooler (The Dow Chemical Company, Midland, Michigan) with ice packs for up to 6 hours and were then stored at -70°C at the clinic for up to 1 month. Periodically, batches of specimens were shipped overnight on dry ice to the main repository for storage at -70°C for 3-36 months. The self-collected specimens were stored at room temperature away from excess heat for up to 48 hours and were then shipped overnight at room temperature to the main repository for storage at -70°C for 3-18 months.

DNA extraction and quantitation

Aliquots (10 ml, unless <10 ml was collected) of the oral rinse specimens were processed for DNA extraction. Samples were thawed at room temperature, were agitated for 1.5 hours at 70°C to inactivate potential pathogens, were incubated with 160 units of ribonuclease for 30 minutes at 65°C, were treated overnight with 80 units of proteinase K at 50°C, and then were heated to 95°C for 30 minutes. DNA was precipitated overnight at -20°C by using sodium acetate (0.1× volume of sample) and isopropanol (1.5× volume of sample plus sodium acetate) supplemented with glycogen (final concentration, 20 µg/ml). Samples were spun at 10,000 rpm for 30 minutes in a Beckman L5-50B swinging bucket ultracentrifuge (Beckman Scientific Instruments, Fullerton, California) at 4°C. DNA was washed with 70 percent ethanol, was resuspended in Tris (hydroxymethyl) aminomethane-chloride (Tris-HCl) and ethylenediaminetetraacetic acid buffer (10

mM Tris-HCl, 1 mM ethylenediaminetetraacetic acid) at a concentration of 0.05 µg/µl, and was stored at 4°C prior to use.

DNA yields were determined spectrophotometrically by using a Beckman DU-640 Spectrophotometer (Beckman Scientific Instruments, Fullerton, California). To ensure accurate readings, sample concentrations were adjusted to achieve optical density values of between 0.1 and 1.0 at a wavelength of 260 λ. To enable the quantity of DNA in the entire oral rinse specimen to be estimated, the quantity of DNA in the aliquot was divided by the proportion of the total sample volume in the aliquot. DNA integrity was assessed on 0.4 percent agarose gels; samples contained high-molecular-weight DNA but showed evidence of some degradation.

PCR-based genotyping assays

DNA quality was assessed by the samples' ability to amplify when the PCR was used (8, 9). A 145-base-pair fragment of the *ADH3* gene was amplified as described previously (6). Subjects' *NAT1* genotypes were determined according to the method of Bell et al. (10): primers N1208F and a five-base-pair extension of primer N1536R (5'-TAACCACAGGCCATCTT-TAGAA) were used, which resulted in a 350-base-pair amplification product. Multiplex PCR was performed for a 332-base-pair fragment of the *CYP1A1* gene, a 1,343-base-pair fragment of the *GSTT1* gene, and a 273-base-pair fragment of the *GSTM1* gene by adding DNA to buffer (100 mM of Tris-HCl (pH 8.3), 15 mM of MgCl₂, and 500 mM of KCl), 200 µM of 2'-deoxynucleoside-3'-triphosphates (Pharmacia, Piscataway, New Jersey), 2.5 units of Taq polymerase (Perkin-Elmer, Norwalk, Connecticut), 0.8 µM of primers for *GSTM1* and 0.4 µM of primers for *CYP1A1* (11), and 1.2 µM of primers 5'-TTCCTTACTGGTCCTCA-CATCTC and 5'-TCACCGATCATGGCCAGCA specific for *GSTT1*. The multiplex PCR consisted of an initial denaturing time of 4 minutes at 94°C, which was followed by 35 cycles of denaturing (1 minute at 95°C), annealing (1 minute at 65°C), and extension (1 minute at 70°C). Then, a final extension step was performed (4 minutes at 70°C). For all PCR-based assays, two reviewers independently read the results blinded to subjects' case-control status, subject characteristics, and method used to collect the sample.

Statistical methods

The mean values of log₁₀-transformed DNA yields among groups were compared by use of the *F* statistic in an analysis of covariance (12), as computed by the PC-SAS (13) computer software procedure PROC GLM. Logarithmic transformation emphasizes differ-

ences at low values over those at high values and was used to evaluate factors that distinguish samples with inadequately low DNA yields from those with higher yields. Nonparametric tests for differences in median DNA yields among groups (PC-SAS procedure, PROC NPAR1WAY) showed similar results and are not presented here. To test for linear trend using PROC GLM, the categorical age variable was treated as a continuous variable in a model in which each level was represented by the median value of that category in the control group. Through the use of PROC GLM, statistical tests of differences by case-control status, age, and gender were adjusted for the other factors under study (i.e., age, gender, collection method, and interviewer); tests of collection method effects were adjusted for age and gender; and tests of interviewer effects among self-collection participants were adjusted for age and gender (among cases) and for gender (among controls, all of whom were aged 65-79 years). All tests of significance were two-tailed.

Chi-square tests, as implemented by the PC-SAS (13) computer software procedure PROC FREQ, were used to assess differences among distributions of categorical variables, including participation status. The EXACT option of PROC FREQ, corresponding to Fisher's exact test (14), was used to compare PCR assay outcomes (success or failure). The Mantel-Haenszel chi-square test (15) within PROC FREQ was used to assess trend in the relation between PCR assay outcome and categories of DNA yield. All tests of significance were two-tailed.

RESULTS

Participation rates among controls

Participation was higher for self-collection (54 of 55 subjects (98.2 percent)) than for clinic collection (94 of 133 subjects (70.7 percent)) ($p = 0.001$). The major reason for nonparticipation was subject refusal (self-collection, $n = 1$; clinic collection, $n = 31$). Participants in the clinic collection were more likely to be older than nonparticipants ($p = 0.05$), but the gender distribution was similar (data not shown).

DNA yields from control samples, by collection method

The quantity of DNA obtained per sample ranged from 2.0 to 204.5 µg (median, 25.9 µg), with similar average yields (mean log₁₀ yields) obtained by self-collection and clinic collection ($p = 0.37$) (table 1). However, samples collected from controls under the direction of interviewer D contained approximately half as much DNA as those directed by the other three

TABLE 1. DNA yields from self-collected and clinic-collected oral rinse samples, Puerto Rico, 1992-1995

	No.	DNA yield (μg)				No. (%) DNA yield $< 5 \mu\text{g}$ †
		Range	Median	Mean (SD)*	Mean log ₁₀ yield (SD)†	
<i>Controls</i>						
Gender						
Male	115	2.0-204.5	27.7	35.9 (33.0)	1.41 (0.36)	3 (2.6)
Female	34	2.8-91.0	13.6	20.8 (19.3)	1.17 (0.36)	2 (5.9)
					$p = 0.001$	$p > 0.32$
Age (years)						
21-44	6	11.6-51.4	23.0	28.4 (16.9)	1.39 (0.27)	0 (0.0)
45-64	41	2.0-89.5	36.1	36.7 (24.4)	1.44 (0.38)	1 (2.4)
65-79	102	2.8-204.5	21.8	31.0 (33.8)	1.32 (0.38)	4 (3.9)
					p for trend = 0.10	$p > 0.99$
Collection method						
Clinic	94	2.0-204.5	26.3	34.3 (35.3)	1.36 (0.39)	3 (3.2)
Self	55	3.5-91.0	24.4	29.2 (21.7)	1.34 (0.35)	2 (3.6)
					$p = 0.37$	$p > 0.99$
Interviewer						
A	10	6.7-77.9	36.6	36.5 (26.6)	1.42 (0.40)	0 (0.0)
B	16	7.0-77.8	29.1	33.2 (20.4)	1.44 (0.30)	0 (0.0)
C	9	11.3-91.0	28.5	37.9 (23.5)	1.51 (0.26)	0 (0.0)
D	20	3.5-75.2	14.3	18.4 (15.6)	1.15 (0.32)	2 (10.0)
					$p = 0.02$	$p = 0.65$
Total	149	2.0-204.5	25.9	32.4 (31.0)	1.36 (0.38)§	5 (3.4)¶
<i>Cases</i>						
Gender						
Male	138	0.9-907.5	25.1	54.8 (101.0)	1.38 (0.56)	14 (10.1)
Female	19	2.4-100.2	17.1	27.0 (26.1)	1.24 (0.44)	2 (10.5)
					$p = 0.29$	$p > 0.99$
Age (years)						
21-44	6	0.9-907.5	10.8	173.2 (361.9)	1.25 (1.08)	2 (33.3)
45-64	72	2.0-540.1	25.4	49.6 (76.0)	1.37 (0.54)	7 (9.7)
65-79	79	1.2-343.9	24.5	43.9 (58.2)	1.36 (0.51)	7 (8.9)
					p for trend = 0.79	$p = 0.19$
Collection method						
Clinic	66	1.2-907.5	27.6	57.4 (120.1)	1.40 (0.55)	6 (9.1)
Self	91	0.9-540.1	18.0	47.1 (73.1)	1.34 (0.55)	10 (11.0)
					$p = 0.47$	$p = 0.79$
Interviewer						
A	37	0.9-203.2	18.0	45.6 (56.1)	1.35 (0.55)	4 (10.8)
B	21	3.4-139.2	17.0	32.4 (38.1)	1.27 (0.46)	2 (9.5)
C	1	99.7	99.7	99.7	2.00	0 (0.0)
D	32	2.0-540.1	19.7	57.0 (102.9)	1.34 (0.60)	4 (12.5)
					$p = 0.55$	$p > 0.99$
Total	157	0.9-907.5	24.5	51.5 (95.5)	1.36 (0.55)§	16 (10.2)¶

* SD, standard deviation.

† Comparison adjusted for other factors in this table, as described in Materials and Methods.

‡ Unadjusted comparison, Fisher's exact test.

§ $p = 0.95$, cases vs. controls.

¶ $p = 0.02$, cases vs. controls.

interviewers (A-C) ($p = 0.02$). Data were not available to investigate possible reasons for inter-interviewer differences. DNA yields were higher among men than

women ($p = 0.001$) but did not differ substantively by subjects' age (p for trend = 0.10) or positive histories of tobacco use or alcohol drinking (data not shown).

Five samples (3.4 percent) contained less than 5 μ g of DNA; the presence of such a low yield was not associated with subjects' age or gender, collection method, or interviewer.

PCR amplification results from control samples, by collection method

The suitability of samples for PCR was assessed with the *ADH3*, *NAT1*, and multiplex *CYP1A1/GSTT1/GSTM1* genotyping assays (table 2). Both collection methods provided high-quality DNA, with similar amplification success rates for self-collected (range, 76.4–98.2 percent) and clinic-collected (range, 87.2–97.9 percent) samples for each of the assays used. PCR success rates were somewhat lower for control samples collected under the direction of interviewer D versus the other interviewers, particularly for the *NAT1* assay ($p = 0.06$). Subjects' genotypes were determined for at least two of the three assays for nearly all of the

self-collected (53 of 55 (96.4 percent)) and clinic collected (92 of 94 (97.9 percent)) samples. Samples with the lowest DNA yields had a greater proportion of amplification failures for at least two of the genotyping assays ($p = 0.015$) (figure 1). Amplification results were not associated with subjects' gender, age, or positive histories of alcohol drinking and tobacco use (data not shown).

Participation rates, DNA yields, and PCR amplification results from case samples

As was true for controls, case participation was higher for self-collection (87 of 91 subjects (95.6 percent)) than for clinic collection (66 of 98 subjects (67.3 percent)) ($p < 0.001$). Reasons for nonparticipation in the clinic collection included dying or becoming too ill in the time interval between interview and specimen collection ($n = 16$) and refusal ($n = 10$), whereas all nonparticipants ($n = 4$) in the self-collection refused.

TABLE 2. Performance of self-collected and clinic-collected oral rinse samples in three PCR*-based genotyping assays, Puerto Rico, 1992–1995

	No.	Successful PCR amplification (no. (%))			
		<i>ADH3</i>	<i>NAT1</i>	<i>CYP1A1/GSTT1/GSTM1</i>	Any 2 assays
<i>Controls</i>					
Total	149	146 (98.0)†	124 (83.2)‡	143 (96.0)§	144 (97.3)¶
Collection method					
Clinic	94	92 (97.9)	82 (87.2)	91 (96.8)	92 (97.9)
Self	55	54 (98.2)	42 (76.4)	52 (94.5)	53 (96.4)
		$p > 0.99$	$p = 0.11$	$p = 0.67$	$p = 0.63$
Interviewer					
A	10	10 (100.0)	7 (70.0)	10 (100.0)	10 (100.0)
B	16	16 (100.0)	14 (87.5)	16 (100.0)	16 (100.0)
C	9	9 (100.0)	9 (100.0)	9 (100.0)	9 (100.0)
D	20	19 (95.0)	12 (60.0)	17 (85.0)	18 (90.0)
		$p > 0.99$	$p = 0.06$	$p = 0.25$	$p = 0.65$
<i>Cases</i>					
Total	157	146 (93.0)†	122 (77.7)‡	140 (89.2)§	140 (89.2)¶
Collection method					
Clinic	66	63 (95.5)	52 (78.8)	60 (90.9)	60 (90.9)
Self	91	83 (91.2)	70 (76.9)	80 (87.9)	80 (87.9)
		$p = 0.36$	$p = 0.85$	$p = 0.61$	$p = 0.61$
Interviewer					
A	37	33 (89.2)	32 (86.5)	34 (91.1)	34 (91.1)
B	21	19 (90.5)	17 (81.0)	19 (90.5)	19 (90.5)
C	1	1 (100.0)	1 (100.0)	1 (100.0)	1 (100.0)
D	32	30 (93.8)	20 (62.5)	26 (81.3)	26 (81.3)
		$p = 0.91$	$p = 0.10$	$p = 0.46$	$p = 0.46$

* PCR, polymerase chain reaction.

† $p = 0.05$, cases vs. controls.

‡ $p = 0.25$, cases vs. controls.

§ $p = 0.03$, cases vs. controls.

¶ $p = 0.006$, cases vs. controls.

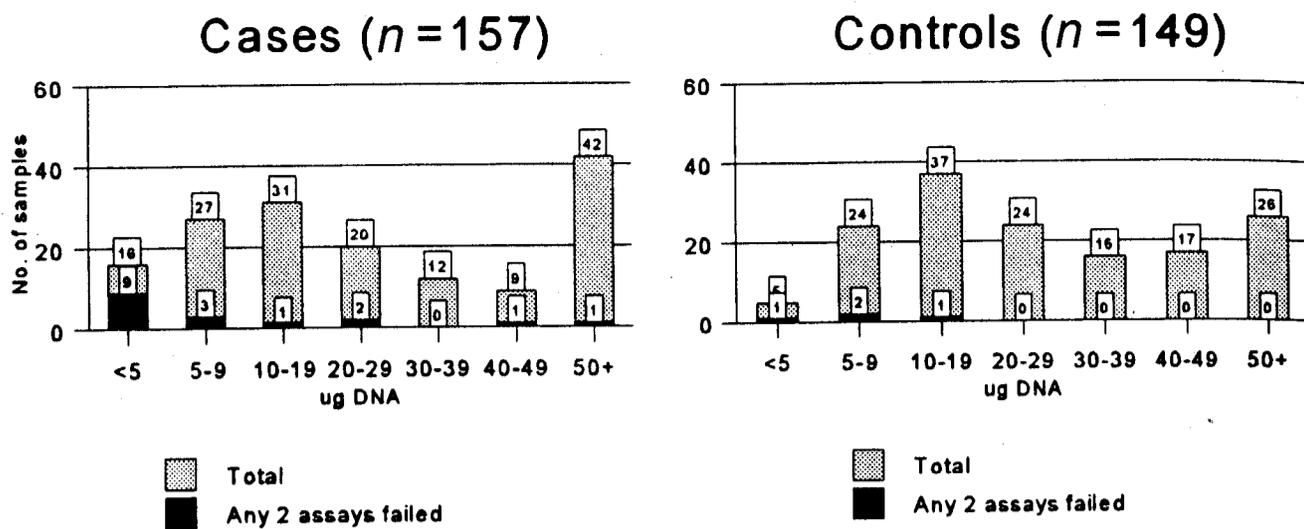


FIGURE 1. Performance of oral epithelial cell samples in three polymerase chain reaction-based genotyping assays (*ADH3*, *NAT1*, multiplex *CYP1A1/GSTT1/GSTM1*). Total numbers of samples and numbers of samples that failed to amplify in at least two of the three assays are shown by DNA yield for controls ($p = 0.015$) and cases ($p = 0.001$), Puerto Rico, 1992–1995.

Subjects who refused the clinic collection were more likely to be female ($p = 0.03$) but did not differ in age from participants (data not shown).

The average DNA yield (mean \log_{10} yield) from cases was similar to that from controls, although a larger proportion of case samples contained less than 5 μg of DNA (10.2 percent) ($p = 0.02$) (table 1). DNA yields were independent of the collection method, interviewer, subjects' age and gender (table 1), and a positive history of tobacco use or alcohol drinking (data not shown).

PCR amplification rates for cases were lower than they were for controls, with 89.2 percent of case samples amplifying in at least two genotyping assays ($p = 0.006$) (table 2). As was observed for control samples, failure to amplify was not associated with gender, age, collection method (table 1), or positive histories of alcohol drinking and tobacco use (data not shown) but was more likely for samples collected under the direction of interviewer D ($p = 0.10$) (table 1) and those with a low DNA content ($p = 0.001$) (figure 1).

DISCUSSION

We found that self-collection of oral epithelial cells under the direction of a trained interviewer yielded similar quantities of DNA as clinic-based collection by a medical technician. Furthermore, over 96 percent of the samples collected from controls by both methods were suitable for PCR-based assays. The DNA yield was lower (controls only) and the proportion of samples successfully amplified by PCR was lower (cases and controls) for samples collected under the direction

of one interviewer, emphasizing the importance of thorough, standardized training. Our results also suggest that short-term storage (up to 3 days) in Standard Transport Medium at room temperature, as was encountered in the field for self-collected samples, does not adversely affect the specimens.

Successful PCR amplification was associated with DNA yield, and the figure 1 data provide some evidence for a threshold (approximately 5 μg) below which PCR is not supported reliably. A greater proportion of samples from cases than from controls contained less than 5 μg of DNA, which may be due to the presence of oral tumor or treatment-related tissue changes. The findings among controls may be generalized to other persons without oral conditions, although some special populations such as the elderly may lack the dexterity necessary to perform the procedure.

Larger quantities of DNA were obtained from men than from women. Men may have larger buccal mucosal surface areas or may brush harder, thereby loosening more cells. However, PCR amplification success rates did not differ by gender, suggesting that for practical purposes, the oral rinse is a suitable method for collecting DNA from subjects of either gender.

It is difficult to directly compare the DNA yields obtained in this study with those reported elsewhere, as different methods for extracting and quantifying DNA were used. DNA yield can also be characterized by the number of PCR-based assays possible per sample. Under the assumption of 25.9 μg of DNA per sample (median yield among controls), we estimate that the samples could support 173 PCR-

based assays by using the same quantity of DNA (0.15 μ g) as was used for each of the genotyping assays in our study. This yield is adequate for epidemiologic studies seeking to investigate several genetic factors.

Participation exceeded 95 percent among subjects eligible to donate oral epithelial cells at their home or workplace at the time of interview, compared with approximately 70 percent among subjects eligible for the full clinic protocol (i.e., blood, urine, and oral epithelial cell collection). Furthermore, when subjects who refused the full biospecimen protocol were given the opportunity to provide only oral epithelial cells, 28 percent did so. Our experience suggests that despite the monetary compensation provided for clinic collection (but not self-collection), subjects preferred self-collection of oral epithelial cell DNA at home or work to clinic collection that included a blood draw and urine specimen.

The observed differences in participation rates may be due to the noninvasiveness of the oral epithelial cell collection protocol, reluctance to undergo a blood draw, or the convenience of home collection at the time of interview. While these are general advantages of the self-collection method, the actual participation rates may vary in other populations because of cultural, logistic, or other reasons. Furthermore, self-collection limits staff exposure to potentially infectious materials and may facilitate large-scale DNA sampling by allowing for collection in geographically remote areas and in situations in which medically trained staff are unavailable because of financial or other constraints. Thus, self-collection of DNA by using oral rinses may be a suitable method for obtaining high-quality samples and achieving high participation rates in epidemiologic studies.

REFERENCES

- Richards B, Skoletsky J, Shuber AP, et al. Multiplex PCR amplification from the *CFTR* gene using DNA prepared from buccal brushes/swabs. *Hum Molec Genet* 1993;2:159-63.
- Hayney MS, Poland GA, Lipsky JJ. A noninvasive "swish and spit" method for collecting nucleated cells for HLA typing by PCR in population studies. *Hum Hered* 1996;46:108-11.
- Lench N, Stanier P, Williamson R. Simple non-invasive method to obtain DNA for gene analysis. *Lancet* 1988;1:1356-8.
- Bugawan TL, Chang JD, Klitz W, et al. PCR/oligonucleotide probe typing of HLA class II alleles in a Filipino population reveals an unusual distribution of HLA haplotypes. *Am J Hum Genet* 1994;54:331-40.
- Aron Y, Swierczewski E, Lockhart A. A simple and rapid micromethod for genomic DNA extraction from jugal epithelial cells. *Allergy* 1994;49:788-90.
- Harty LC, Caporaso NE, Hayes RB, et al. Alcohol dehydrogenase 3 genotype and risk of oral cavity and pharyngeal cancers. *J Natl Cancer Inst* 1997;89:1698-705.
- Hayes RB, Bravo-Otero E, Kleinman DV, et al. Tobacco and alcohol use and oral cancer in Puerto Rico. *Cancer Causes Control* 1999;10:27-33.
- Saiki RK, Scharf S, Faloona F, et al. Enzymatic amplification of β -globin genomic sequences and restriction sites analysis for diagnosis of sickle cell anemia. *Science* 1985;230:1350-4.
- Saiki RK, Gelfand DH, Stoffel S, et al. Primer-directed enzymatic amplification of DNA with a thermostable DNA polymerase. *Science* 1988;239:487-91.
- Bell DA, Stephens EA, Castranio T, et al. Polyadenylation polymorphism in the acetyltransferase 1 gene (*NAT1*) increases risk of colorectal cancer. *Cancer Res* 1995;55:3537-42.
- Ambrosone CB, Freudenheim JL, Graham S, et al. Cytochrome *P450IA1* and glutathione *S*-transferase (M1) genetic polymorphisms and postmenopausal breast cancer risk. *Cancer Res* 1995;55:3483-5.
- Snedecor GW, Cochran WG. *Statistical methods*. 7th ed. Ames, IA: The Iowa State University Press, 1980:365-92.
- SAS Institute, Inc. *The SAS system for Windows*, release 6.12. Cary, NC: SAS Institute, Inc, 1996.
- Kendall M, Stuart A. *The advanced theory of statistics*. Vol 2. New York, NY: Macmillan Publishing Co, 1979:580-5.
- Mantel N, Haenszel W. Statistical aspects of the analysis of data from retrospective studies of disease. *J Natl Cancer Inst* 1959;22:719-48.