Overview and quality assurance for the oral health component of the National Health and Nutrition Examination Survey (NHANES), 2005-08

Bruce A. Dye, DDS, MPH[1]; Laurie K. Barker, MSPH[2], Xiafen Li, MS[1], Brenda G. Lewis, MPH[1]; Eugenio D. Beltrán-Aguilar, DMD, MPH, MS, DrPH[2]

1 Centers for Disease Control and Prevention/National Center for Health Statistics at Hyattsville, MD
2 Centers for Disease Control and Prevention/National Center for Chronic Disease Prevention and Health Promotion/Division of Oral Health at Atlanta, GA

Abstract

The oral health component for the National Health and Nutrition Examination Survey (NHANES) was changed in 2005 from an examination conducted by dentists to an oral health screening conducted by health technologists rather than dental professionals. The oral health screening included a person-based assessment for dental caries, restorations, and sealants. This report provides oral health content information and presents results of data quality analyses that include dental examiner reliability statistics for data collected during NHANES 2005-08. Oral health data are available on 15,342 persons aged 5 years and older representing the civilian, noninstitutionalized population of the United States who participated in NHANES 2005-08. Overall, interrater reliability findings indicate that health technologist performance was excellent with concordance between examination teams and the survey reference examiner being almost perfect for a number of assessments. Concordance for dental caries and sealants (kappa statistics) between health technologists and the survey reference examiner ranged from 0.82 to 0.90 for the combined 4-year period. These findings support the use of health technologists in the assessment of person-based estimators of dental caries and sealant prevalence as part of an oral health surveillance system.

Introduction

Since 1999, the National Health and Nutrition Examination Survey (NHANES) has operated as a continuous, annual survey that provides nationally representative data in 2-year cycles. Like previous national health examination surveys, participants are interviewed in their homes and then complete a health examination at a Mobile Examination Center (MEC). The MEC health examination teams are comprised of specially trained health professionals and support staff. Prior to 2005, the examination teams had licensed dentists conducting the oral health examinations. Beginning in 2005, NHANES switched from a dentist-based oral health examination protocol to an oral health screening conducted by health technologists rather than dental professionals.

In the mid 1990s, the Division of Oral Health (DOH) at the Centers for Disease Control and Prevention (CDC) developed, tested, and validated a simplified data collection method, based on the visual assessment of the oral cavity to provide an alternative to costly surface-based examinations. This tool was based on the concept of medical screening and was originally validated using registered dental hygienist and school nurses as screeners (1). This approach was previously used in the assessment of the prevalence of dental caries

Ethics statement: All study participants gave informed consent in accordance with the Ethics Review Board and study ethic guidelines at the Centers for Disease Control and Prevention. The authors do not have any financial or other competing interests to declare.

Table 1 Sampling Design Characteristics for NHANES, 2005-08

Characteristic	NHANES 2005-06	NHANES 2007-08
Age of the target population	From birth	From birth
Dental exam age	5 years and older	5 years and older
Number of survey locations	30	30
Eligible geographical area		
For sample	50 states + DC	50 states + DC
Groups target for oversampling	Persons 12-19 years and 60 years and older; Mexican-Americans; non-Hispanic Blacks	Persons 60 years and older; all Hispanics; non-Hispanic Blacks
All ages examined in the MEC	9,950	9,762
All persons 5 years and older		
MEC examined	8,305	8,311
All persons 5 years and older		
Oral health examined	7,687	7,655

NHANES, National Health and Nutrition Examination Survey; DC, District of Columbia; MEC, Mobile Examination Center.

among preschool and school-aged children in Oregon and school-aged children in the State of Washington (2,3); and later in the assessment of dental caries in preschool children in Maine, school-aged children in Louisiana, persons with disabilities (including children and adults) participating in Special Olympics programs, and preschool children in Washington State (4-6).

In 1989, an amendment to the Title V Maternal Child Health (MCH) Block Grant legislation (OBRA 89, PL 101-239) required that dental data collection, needs assessment and planning be incorporated into state MCH grant applications. To address this requirement, Health Resources and Services Administration (HRSA) designated the *Healthy People 2000* objective to increase dental sealant prevalence (7) as a required element in state applications for MCH Block Grant funding. In the 1990s, HRSA provided funding for the development of an oral health needs assessment tool to be used by state and local dental public health programs. Subsequently, the Association of State and Territorial Dental Directors formalized and standardized the simplified screening technique, under the label "Basic Screening Survey," to allow states to obtain data for caries experience, untreated decay, and dental sealants (8). These data, in turn, are reported to the CDC National Oral Health Surveillance System (http://www.cdc.gov/nohss/).

NHANES is the only source of data for surveillance of dental diseases and conditions at the national level. To maintain ongoing surveillance for selected oral health outcomes in the United States after 2004, the CDC slightly modified the "Basic Screening Survey" and introduced a Basic Screening Exam (BSE) that could be conducted by nondental professionals for NHANES starting in 2005. A developmental requirement for the BSE was retaining key caries-specific detection protocols from earlier NHANES caries assessments that would permit comparison of data at the person-level. The aim of this article was to describe the 2005-08 NHANES oral health component, the actions taken to ensure data

quality, and the results of analyses assessing data reliability. Information in this article will be useful for researchers in providing them with guidance in properly using the NHANES 2005-08 data for future research as well as for the broader dental public health community by identifying relevant assessments for oral health surveillance.

Methods

NHANES 2005-08 overview

NHANES is conducted by the CDC's National Center for Health Statistics (NCHS) and uses a stratified, multistage probability sampling design to select participants living in the 50 states and the District of Columbia. For NHANES 2005-08, the survey oversampled selected subgroups including non-Hispanic blacks and persons aged 60 years and older. However, there were some oversampling changes from 2005-06 to 2007-08. Mexican-Americans were oversampled in 2005-06, but all Hispanics were oversampled in 2007-08. Additionally, persons aged 12-19 years were oversampled in 2005-06, but not in 2007-08. Sample design characteristics for NHANES 2005-08 are presented in Table 1. Informed consent is obtained for all participants and all data collection protocols are approved by the CDC/NCHS Ethics Review Board. Information on NHANES background, content, analytical guidelines, and data access can be found at: http://www.cdc.gov/nchs/nhanes.htm.

NHANES 2005-08 oral health component

The 2005-08 oral health component included seven questions on oral health quality of life that were first introduced in NHANES 2003-04, and an additional question asking participants to rate the condition of their teeth and gums that had been administered in previous NHANES cycles as well. All these questions were administered during the home

interview. Following completion of the home interview, participants received an MEC appointment to complete a number of clinical and laboratory assessments.

For 2005-08, the MEC oral health assessments were conducted by trained health technologists. NHANES has a long history of successfully using health technologists to administer a variety of examinations including collecting anthropometric (body measurements), audiometric, vision, fitness, and mobility data. Health technologists usually have prior experience and training in the allied health sciences and some are also certified radiology technicians. These individuals also operate the DEXA radiology units which are used to collect data on bone density and osteoporosis. A MEC examination team typically has four health technologists and NHANES has two full-time MEC teams collecting data at the same time. All health technologists were trained and calibrated by the survey reference examiner to conduct the oral health screening. These screenings took place in a designated room inside the MEC that includes a portable dental chair, light, and compressed air.

The first oral health assessment administered by the health technologists in the MEC was an inquiry into denture use and wear. This brief questionnaire was administered to all participants aged 25 years and older. A simplified examination followed that included two elements: a) a tooth count to identify the presence or absence of permanent and/or primary teeth, including retained dental root tips and dental implants; and b) the BSE. The BSE assessed the presence of at least one tooth affected by untreated dental decay, a dental restoration, or a dental sealant. All participants aged 5 years and older were eligible for both the tooth count and BSE assessments. Once the BSE was completed, the MEC oral health screening concluded with an assessment of functional occlusal contacts in those aged 25 years and older. The functional contacts assessment, which evaluated the number and type of contacts between opposing teeth in the maxillary and mandibular arches, was introduced in NHANES 2003-04 and was not changed in the transition from a dentist-based exam to a health technologist screening exam.

The oral health screening was conducted using a dental light for illumination and a disposable mirror. The examining health technologists had access to compressed air to assist in clearing the dental viewing area of residual food debris. The BSE was a fast and simple process by which an examiner visually inspected the oral cavity to detect the presence or absence of specific oral conditions. The assessment procedure began with the central incisor in the upper right quadrant. Each tooth in that quadrant was assessed visually for the specific condition of interest. Inspection systematically proceeded toward the posterior until the second molar had been inspected. Then, the examiner proceeded to the upper left central incisor and proceeded with inspection toward the upper left second molar. The lower left and lower right quadrants followed in sequence.

A full intraoral assessment for one disease/condition was known as a cycle. After the condition was encountered, the ascertainment criterion was fulfilled, and there was no need to complete the cycle for the same condition elsewhere in the mouth. This process is called stop-after-first-encounter (SAFE). For the 2005-08 BSE, the examiner was required to complete three cycles: to ascertain the presence of untreated carious lesions, dental restorations, and pit-and-fissure sealants. The detection criteria for each of the conditions assessed were the same used in previous NHANES (9,10). For example, a dental lesion was classified as "untreated" only if the carious lesion was considered to be cavitated, and a sealant was considered present even if part of the sealant was not visible. Third molars were excluded from the BSE and if the sampled participant was edentulous, the BSE assessment was not performed.

Quality assurance

All data collected in NHANES are directly recorded onto electronic data collection forms to reduce data entry errors. Likewise, automated data management utilities are also used that check entered data for out of range values to reduce errors. The collection of quality oral health data was facilitated by requiring all health technologists to participate in a comprehensive training and calibration period followed by periodic monitoring and recalibration. During the training session, health technologists received instruction from the survey reference examiner that included a slide presentation on the study protocols, a review of dental anatomy, detection criteria, and data entry. Safety and infection control for dental practices were reviewed as well.

Demonstration examinations were conducted by the reference examiner followed by practice (standardization) sessions. During the standardizing phase, the reference examiner and health technologists examined the same set of volunteers. Trainees were encouraged to ask questions regarding the assessment criteria and feedback was provided by the reference examiner after each examination round to systematically minimize differences in the examination findings. A preliminary calibration session followed standardization. Health technologists conducted independent replicate examinations without discussion and the data were analyzed to assess consistency between each trainee and the reference examiner. Health technologists received training and preliminary calibration over a 3-day period in the Washington, DC, metropolitan area. A final calibration session was conducted at the MEC and was performed during normal field operations over a 3- to 4-day period.

Overall, the general framework for training and evaluation of health technologists conducting the oral health exam during 2005-08 was very similar compared with procedures implemented during 1999-2004 when dentists conducted

the NHANES oral health examinations (9,10). The reference examiner visited each MEC health technologist team two to three times each year to observe data collection and to randomly replicate approximately 25 to 30 oral health examinations during each visit. Data from these replicate exams were used to produce interrater reliability statistics to objectively evaluate examiner performance. Although health technologists were aware of the interrater evaluations being conducted, examiners were blinded to each other's observations.

For this report, reliability statistics produced were limited to percent agreement and kappa statistics using Statistical Analysis Software (version 9.2, SAS Institute Inc., Cary, NC, USA). Kappa statistics are presented only if the number of observations was >1 for the concordant cells. Observations with values "cannot assess" were excluded from the calculations. For the evaluation of tooth count, Kappa statistics were produced for four assessments. These were complete tooth loss, retention of all third molars, having at least one retained root tip, and tooth retention, which was the total number of primary and permanent teeth present. Kappa statistics calculated for the BSE included having at least one tooth with dental caries, one tooth with a dental restoration, and one tooth with a dental sealant. Five different posterior functional contact summary variables were used to assess interrater reliability. These included the presence of five or fewer total posterior contacts, and the presence of at least one contact in the premolar and molar zones for both the right and left sides. This construct for functional contacts has been presented before and the reader can compare Kappa statistics from this report with an earlier report (9) because the reference examiner did not change between the different data collection periods.

Because each MEC team is staffed with four health technologists, and each technologist is randomly assigned to conduct an oral health exam, it was not possible to administer a sufficient number of replicate exams on each examiner to allow for individual interrater reliability calculations. Moreover, not all health technologists remained on the survey for the entire 4 years and some were reassigned each year to a different team to minimize team bias. Consequently, interrater reliability statistics for this report represent the overall team of health technologists.

Results

The number of individuals participating in the home interview, MEC examination, and the oral health examination in NHANES 2005-08 is shown in Table 2. There were 7,687 sampled persons participating in the oral health exam in 2005-06 and 7,655 participating in the 2007-08 oral health exam. For 2005-06, the proportion of participants living in households at or below the Federal Poverty Level (FPL)

having an oral health exam was 22.5 percent whereas the proportion of participants living in households at least twice as high as the FPL having an oral health exam was 52.1 percent. The proportions were similar for these two groups in 2007-08 (24.2 percent and 48.3 percent). For Mexican-Americans, the proportion having an oral health exam was 25.7 percent in 2005-06 but was 19.8 percent in 2007-08. For non-Hispanic blacks, the proportion having an oral health exam was 26.6 percent and 22.0 percent during 2005-06 and 2007-08, respectively. During the same two time periods, the proportion of non-Hispanic whites having an oral health exam was 40.0 percent and 42.5 percent. For persons living in households below the FPL, not completing high school, and current smokers, the proportions of individuals completing a home interview and the oral health examination was similar during the two data collection cycles.

Table 3 shows the interrater reliability statistics for each of the 2-year data collection cycles and for the period 2005-08 combined. For the period 2005-06, the kappa statistics for the tooth count assessment ranged from 1.00 to 0.52 with a percent agreement ranging from 100 percent to 96 percent. With the exception of a third molar assessment from a single team (0.52), the kappa scores ranged from 0.89 to 1.00 for the tooth count assessments. Both teams scored a 1.00 for assessing edentulism. Kappa scores ranged from 0.93 to 1.00 for tooth retention and retained tooth roots. For the BSE, kappa scores were 0.80 and 0.78 for untreated caries, 0.92 and 0.83 for dental restorations, and 0.79 and 0.85 for dental sealants. Kappa scores ranged from 0.86 to 1.00 for functional contacts for both teams.

For 2007-08, the kappa statistics for the tooth count assessments ranged from 0.79 to 1.00 with a percent agreement ranging from 100 percent to 97 percent. Both teams scored a 1.00 for assessing edentulism, and for tooth retention and retained root tips, the kappa scores ranged from 0.87 to 0.98. The kappa statistics for the BSE were very similar for both teams. Scores were 0.86 and 0.84 for untreated caries, 0.92 and 0.87 for dental restorations, and 0.90 and 0.89 for dental sealants. The kappa scores for functional contacts ranged from 0.81 to 1.00.

When reviewing the combined 2005-08 period, the kappa statistics for tooth count, excluding the third molar assessment, ranged from 0.89 to 1.00. For the third molar assessment, the kappa scores were 0.90 and 0.69 with a basic concordance ranging from 97-98 percent agreement. For the BSE assessment, kappa scores were 0.83 and 0.82 for untreated caries, 0.90 and 0.88 for dental restorations, and 0.87 and 0.83 for dental sealants. Functional contact kappa scores ranged from 0.85 to 0.98.

Interrater reliability statistics assessing dental caries, restorations, and sealants for the mixed dentition is presented in Table 4. Overall, the kappa scores were 0.87 and 0.76 for untreated caries, 0.91 for dental restorations, and 1.00 and

Table 2 Number of Sampled Persons Aged 5 Years and Older with Interview, MEC, and Oral Health Exams by Selected Demographic Characteristics, NHANES 2005-08

Characteristic	NHANES 2005-06			NHANES 2007-08		
	HIQ	MEC	OHX	HIQ	MEC	OHX
Age						
5-11 years	1,375	1,325	1,295	1,452	1,394	1,363
12-17 years	1,704	1,646	1,544	945	922	864
18-24 years	1,120	1,075	967	752	734	668
25-34 years	977	936	853	923	887	786
35-44 years	846	809	735	999	966	862
45-54 years	781	766	703	1,000	977	889
55-64 years	650	623	575	998	963	875
65-74 years	591	569	533	789	764	708
75+ years	598	556	482	767	704	640
Sex						
Male	4,202	4,042	3,743	4,299	4,145	3,856
Female	4,440	4,263	3,944	4,326	4,166	3,799
Race and ethnicity						
Mexican-American	2,207	2,123	1,976	1,711	1,636	1,517
Other Hispanic	268	252	234	1,017	975	883
Non-Hispanic White	3,442	3,306	3,076	3,594	3,476	3,255
Non-Hispanic Black	2,316	2,238	2,043	1,928	1,867	1,684
Other race	409	386	358	375	357	316
Poverty status						
Less than 100% FPL	1,831	1,787	1,656	1,915	1,860	1,702
100%-199% FPL	2,095	2,016	1,868	2,163	2,084	1,927
200% or higher FPL	4,281	4,107	3,835	3,777	3,642	3,397
Education						
Less than high school	1,239	1,181	1,051	1,740	1,649	1,461
High school	1,041	1,005	919	1,328	1,285	1,166
More than high school	2,154	2,066	1,909	2,401	2,321	2,129
Smoking history						
Current smoker	942	899	819	1,195	1,153	1,051
Former smoker	1,191	1,141	1,056	1,443	1,386	1,278
Never smoker	2,304	2,215	2,004	2,831	2,716	2,427
Total	8,642	8,305	7,687	8,625	8,311	7,655

Notes: Education and Smoking are only for sampled persons aged 25 years and older.
NHANES, National Health and Nutrition Examination Survey; HIQ, Home Interview; MEC, Mobile Examination Center; OHX, Oral Health Examination; FPL, Federal Poverty Level.

0.86 for dental sealants during the 2005-08 survey period for children aged 5-12 years. However, the difference in kappa scores for untreated caries between Teams 1 and 2 was large during 2005-06. Although the percent agreement for Team 1 was 87.5 percent and 95 percent for Team 2, the kappa scores were 0.45 and 0.89 for Teams 1 and 2, respectively.

Discussion

NHANES is the only source of data for surveillance of dental diseases at the national level. The NHANES 2005-08 oral health screening provided comparable data estimators generated from data previously collected in national surveys, because it used the same diagnostic criteria. Moreover, the BSE provided data that are comparable with those collected at

state and local levels using the BSS protocol. Before implementation in the NHANES protocol, the BSE was piloted in 2004, focusing on three important aims: a) to produce a valid oral health exam that could be conducted in 2-3 minutes; b) to have health technologists reliably conduct the exam; and c) to have a minimally invasive screening exam that would be accepted by the US population.

Earlier development of the BSE component of the NHANES oral health screening could be traced back to 1994 when the CDC-DOH conducted a methodological assessment of an intraoral screening protocol for assessing oral health status among schoolchildren in rural Georgia (1). During this study, a registered nurse or a dental hygienist ascertained the presence/absence of seven dental conditions among 632 children. Results from these assessments were

Table 3 Interrater Reliability Statistics for Selected Oral Health Conditions, NHANES 2005-08

#	Assessments	NHANES 2005-06				NHANES 2007-08				NHANES 2005-08			
		n	%	Kappa	ASE	*n*	%	kappa	ASE	*n*	%	kappa	ASE
	Tooth count												
1	Edentulism	152	100.00	1.00	0.00	109	100.00	1.00	0.00	261	100.00	1.00	0.00
2	Edentulism	112	100.00	1.00	0.00	156	100.00	1.00	0.00	268	100.00	1.00	0.00
1	Tooth retention	152	99.34	0.99	0.01	109	99.08	0.98	0.02	261	99.23	0.98	0.01
2	Tooth retention	112	97.32	0.94	0.03	156	96.80	0.92	0.03	268	97.02	0.93	0.02
1	One or more retained root tips	152	99.34	0.93	0.07	109	100.00	1.00	0.00	261	99.62	0.96	0.04
2	One or more retained root tips	112	100.00	1.00	0.00	156	98.08	0.87	0.07	268	98.88	0.89	0.06
1	Have all four third molars	152	98.03	0.89	0.07	109	99.08	0.94	0.06	261	98.47	0.90	0.05
2	Have all four third molars	112	95.54	0.52	0.19	156	97.44	0.79	0.10	268	96.64	0.69	0.10
	BSE assessment												
1	Has untreated caries	152	92.11	0.80	0.05	109	94.50	0.86	0.05	261	93.10	0.83	0.04
2	Has untreated caries	112	91.07	0.78	0.07	156	93.59	0.84	0.05	268	92.54	0.82	0.04
1	Has restorations	152	96.05	0.92	0.03	109	93.58	0.87	0.05	261	95.02	0.90	0.03
2	Has restorations	112	91.96	0.83	0.05	156	96.15	0.92	0.03	268	94.40	0.88	0.03
1	Has dental sealants	152	95.40	0.79	0.07	109	98.17	0.90	0.07	261	96.55	0.83	0.05
2	Has dental sealants	112	96.43	0.85	0.07	156	98.08	0.89	0.06	268	97.39	0.87	0.05
	Functional contacts												
1	Five or less posterior contacts	70	98.57	0.96	0.04	52	96.15	0.88	0.08	122	97.54	0.93	0.04
2	Five or less posterior contacts	73	98.63	0.93	0.07	104	92.31	0.81	0.06	177	94.92	0.85	0.05
1	Contact in right molar zone	70	98.57	0.96	0.04	52	100.00	1.00	0.00	122	99.18	0.98	0.02
2	Contact in right molar zone	73	98.63	0.88	0.12	104	96.15	0.90	0.05	177	97.18	0.90	0.04
1	Contact in right premolar zone	70	100.00	1.00	0.00	52	98.08	0.91	0.09	122	99.18	0.97	0.03
2	Contact in right premolar zone	73	98.63	0.90	0.10	104	94.23	0.83	0.07	177	96.05	0.85	0.06
1	Contact in left molar zone	70	98.57	0.96	0.04	52	98.08	0.94	0.06	122	98.36	0.95	0.03
2	Contact in left molar zone	73	97.26	0.86	0.10	104	96.15	0.90	0.05	177	96.61	0.90	0.04
1	Contact in left premolar zone	70	98.57	0.95	0.05	52	100.00	1.00	0.00	122	99.18	0.97	0.03
2	Contact in left premolar zone	73	98.63	0.88	0.12	104	96.15	0.87	0.07	177	97.18	0.87	0.06

NHANES, National Health and Nutrition Examination Survey; *n*, number of replicate exams; %, percent agreement; ASE, asymptotic standard error (kappa); #, health technologist examination team number.

compared against prevalence figures obtained in the same children by a trained oral epidemiologist using the standard visual-tactile protocol used in prior US national surveys. Validity was assessed in terms of sensitivity, specificity, and predictive values, and reliability was assessed by percent agreement and kappa on duplicate examinations. The findings indicated that screenings were valid intraoral assessments for dental caries: all values were higher than 90 percent and no differences were detected between the two screeners (nurse and hygienist) and the reference examiner. Although sensitivity for prevalence of dental sealants was lower than for dental caries, dental sealant data was determined to be reliable with the likelihood of underreporting the prevalence of dental sealants for surveillance activities.

This report is part three of a series of methodology papers that now cover ten years of oral health data

Table 4 Interrater Reliability Statistics for Basic Screening Exam for Children Aged 5-12 Years, NHANES 2005-08

#	BSE Assessment	NHANES 2005-06				NHANES 2007-08				NHANES 2005-08			
		n	%	Kappa	ASE	*n*	%	kappa	ASE	*n*	%	kappa	ASE
1	Has untreated caries	32	87.50	0.45	0.22	28	100.00	1.00	0.00	60	93.33	0.76	0.11
2	Has untreated caries	20	95.00	0.89	0.10	31	93.55	0.84	0.11	51	94.12	0.87	0.08
1	Has restorations	32	93.75	0.86	0.10	28	100.00	1.00	0.00	60	96.67	0.91	0.06
2	Has restorations	20	90.00	0.69	0.21	31	100.00	1.00	0.00	51	96.08	0.91	0.06
1	Has dental sealants	32	90.63	0.71	0.15	28	100.00	1.00	0.00	60	95.00	0.86	0.07
2	Has dental sealants	20	100.00	1.00	0.00	31	100.00	1.00	0.00	51	100.00	1.00	0.00

NHANES, National Health and Nutrition Examination Survey; *n*, number of replicate exams; %, percent agreement; ASE, asymptotic standard error (kappa); #, health technologist examination team number.

collection on the continuous NHANES (9,10). In this report, a broad range of data quality issues for the NHANES 2005-08 oral health component are presented including measures of examiner reliability. Although researchers may use different standards in ascertaining the strength of agreement between examiners, for this report we relied on a widely used guideline proposed by Landis and Koch for interpreting kappa scores (11). A kappa statistic ≤ 0 is reflective of having "poor agreement," >0 but ≤ 0.20 is "slight agreement," 0.21-0.40 is "fair agreement," 0.41-0.60 is "moderate agreement," 0.61-0.80 is "substantial agreement," and >0.80 is "almost perfect agreement,"

During NHANES 2005-08, the interrater reliability statistics indicated that team agreement with the reference examiner was almost perfect for the tooth count assessment when excluding assessment for third molars. There was a difference in the measure of agreement between Teams 1 and 2 regarding third molar assessments between 2005-06 (moderate versus almost perfect) and 2007-08 (substantial versus almost perfect). Assessing for third molars can be problematic when individuals are partially edentulous and tooth morphology may not be helpful in distinguishing between second molars and those third molars that may have drifted forward in the absence of second molars. Nevertheless, the health technologists were able to produce substantial agreement with the reference examiner (a dentist) over the 4-year period regarding proper identification of the presence/absence of third molars. Properly coding third molars is important in NHANES because these teeth are typically excluded from caries and edentulism assessments.

Agreement was considered almost perfect for all three BSE assessments during 2005-08. In 2005-06, the kappa statistics for both teams for untreated caries, dental restorations, and dental sealants were in the upper 0.70s or higher, which would classify both teams as having substantial agreement with the reference examiner. Additionally, the kappa scores increased for both teams from 2005-06 to 2007-08 for untreated caries and dental sealants. Although only a marginal improvement in the kappa statistics, this increase in agreement may represent a slight gain in the overall team learning curve as more experienced health technologists, who conducted the exam over a 3- to 4-year period, more favorably affected the overall team performance compared with less experienced health technologists, who only conducted the exam for 1 to 2 years. Finally, agreement was considered almost perfect for the functional contacts assessment and this was comparable with interrater reliability findings observed for dentists during NHANES 2003-04 (8).

Overall, agreement was almost perfect for assessing untreated caries, dental restorations, and dental sealants in the mixed dentition during 2005-08. Although there were substantial differences between the two teams during 2005-06 regarding assessing untreated caries among children aged 5-12

years, concordance among Team 1 members with the reference examiner improved from moderate agreement to perfect agreement during 2007-08. This outcome may be explained by three factors. First, the variability associated with producing the 2005-06 untreated caries kappa statistic for Team 1 was considerably high (0.22) yet that the percent agreement was 87.5 percent. This could indicate that the number of discordant observations was much lower in one category compared with the other. Given the sample size of the repeat exams for this age group, this could indicate that there was either under- or over-calling of the presence of untreated caries. Indeed, there were four events which team members did not call untreated caries whereas the reference examiner did and there were no events where team members called untreated caries and the reference examiner did not. Therefore, the disagreement observed during 2005-06 between Team 1 and the reference examiner indicates bias toward under-calling the prevalence of untreated caries in the mixed dentition. The two factors that most likely impacted the improvement in kappa scores from 2005-06 to 2007-08 was the redistribution of some of the team members that routinely occurs between 2-year survey periods and the ongoing quality assurance process, which included periodic field visits.

Examiner reliability findings from NHANES 2005-08 confirm earlier suggestions that training nondental health workers may be an accurate and less resource-intense process for obtaining basic oral health data (12). In an earlier study, Warren *et al.* found that trained nurses were able to reproduce a tooth count exam administered by dentists in 86 percent of the subject exams conducted. However, there were some limitations in this study that does affect the interpretation of findings. Training and calibration was conducted on a single day, and there was no follow-up regarding the evaluation of inter- or intra-examiner reliability during the length of the main study, which required only intermittent oral health data collection.

Although 2-year data cycles from NHANES 1999-2000 and forward are considered nationally representative and general assessments of oral health data reliability for NHANES 2005-08 are considered very good, analysts should be cautious in comparing 2-year estimates from the continuous NHANES because of three important considerations. In a prior report discussing data quality and comparability between the two 3-year survey phases from NHANES 1988-94, the authors described a variety of survey design issues, including sampling variation that could influence oral health prevalence differences between the two phases (1988-92 and 1992-94) even though both phases were nationally representative and examiner reliability was considered to be good (13). The second important issue regarding the use of 2-year data cycles is related to using insufficient sample size to calculate oral health estimates, which could produce statistically unreliable estimates. To ensure reliability of estimates,

analysts should routinely evaluate whether the denominator count is >30 and the relative standard error is greater than 30 percent. For most oral health calculations, using 4 years of data for analyses will reduce the effect of sampling variation between the two 2-year periods and produce more accurate estimates. The third consideration affecting interpretation and comparability of oral health data between 2-year data cycles is related to the possibility of measurement bias as a result in changes in data collection methodology between 1999-2004 and 2005-08.

NHANES oral health data collection has changed again. For NHANES 2009-10, oral health data are being collected by registered dental hygienists using the person-based BSE and a tooth-based full-mouth periodontal examination. The implementation of a full-mouth periodontal exam represents another significant milestone for oral health surveillance and dental public health research in the United States. For the first time in national health examination survey history, six probing measures per tooth will be made. This expanded periodontal assessment will permit the full application of a recently developed definition of periodontitis by the CDC and American Academy of Periodontology for use in surveillance and epidemiologic research (14).

The oral health data collected during NHANES 2005-08 was the first oral health examination data collected in a national US examination survey that was obtained by health technologists rather than dental professionals. Oral health data were collected by MEC health technologists using person-based assessments for dental caries, restorations, and sealants. Tooth-based assessments were also conducted to assess the presence/and absence of individual teeth and functional occlusal contacts. For this 4-year period, overall inter-rater reliability findings indicate that health technologist performance was excellent with concordance between examination teams and the survey reference examiner being almost perfect for a number of assessments, including the BSE. Intensive training, monitoring and follow-up was an important factor in producing quality oral health data, but an equally important factor may be having the trained health technologists collect the data under routine or "full-time" conditions. The data presented in this article indicate that the oral health screening exam, which included the BSE assessment, can produce reliable estimates when conducted by medical technologists with the appropriate level of training and data collection oversight.

Acknowledgments

References

1. Beltrán ED, Malvitz DM, Eklund SA. Validity of two methods for assessing oral health status in populations. *J Public Health Dent.* 1997;**57**:206-14.

2. Phipps KR, Mason JD. *Oral health needs assessment 1991-93*. Portland (OR): Oregon Department of Human Resources, Child Health Section; 1993.

3. Phipps K, Hines B, Scheer M. *Washington state smile survey. A children's oral health assessment report*. Olympia, WA: Maternal and Child Health, Oral Health Program; 1996.

4. Sutherland JN, Romaguera R. Oral health needs among children attending Louisiana public schools with health centers: phase I. *La Dent J.* 1997;**56**:14-6.

5. White JA, Beltrán ED, Malvitz DM, Perlman SP. Oral health status of special athletes in the San Francisco Bay Area. *J Calif Dent Assoc.* 1998;**26**:347-54.

6. Phipps K, Beltran E, Hines B. ECC in infants and toddlers: a statewide perspective. [Abstract]. 64th Annual Meeting of the American Association of Public Health Dentistry. Portland, Oregon, 2001.

7. U.S. Department of Health and Human Services. *Healthy people 2000. National health promotion and disease prevention objectives*. Washington, DC: U.S. Government Printing Office; 1990.

8. Association of State and Territorial Dental Directors. *Basic screening surveys: an approach to monitoring community oral health*. Columbus (OH): Association of State and Territorial Dental Directors; 1999.

9. Dye BA, Nowjack-Raymer R, Barker LK, Nunn JH, Steele JG, Tan S, Lewis BG, Beltran ED. Overview and quality assurance for the oral health component of the National Health and Nutrition Examination Survey (NHANES), 2003-2004. *J Public Health Dent.* 2008;**68**:218-26.

10. Dye BA, Barker LK, Selwitz RH, Lewis BG, Wu T, Fryar CD, Ostchega Y, Beltran ED, Ley E. Overview and quality assurance for the national health and nutrition examination survey (NHANES) oral health component, 1999-2002. *Community Dent Oral Epidemiol.* 2007;**35**:140-51.

11. Landis JR, Koch GG. The measurement of observer agreement for categorical data. *Biometrics.* 1977;**33**:159-74.

12. Warren JJ, Levy SM, Hand JS. The accuracy of tooth loss data collected by nurses. *Spec Care Dentist.* 1999;**19**:75-8.

13. Winn DM, Johnson CL, Kingman A. Periodontal disease estimates in NHANES III: clinical measurement and complex sample design issues. *J Public Health Dent.* 1999;**59**:73-8.

14. Page RC, Eke PI. Case definitions for use in population-based surveillance of periodontitis. *J Periodontol.* 2007;**78**:1387-99.