

Overview and quality assurance for the National Health and Nutrition Examination Survey (NHANES) oral health component, 1999–2002

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: 140–151. © Blackwell Munksgaard, 2007

Abstract – The Oral Health Component of the 1999–2002 National Health and Nutrition Examination Survey (NHANES) is a collaborative effort between the National Institute of Dental and Craniofacial Research (NIDCR), the National Center for Chronic Disease Prevention and Health Promotion, Division of Oral Health (NCCDPHP/DOH), and the National Center for Health Statistics (NCHS). The current NHANES is designed as a continuous survey with data released on a 2-year cycle to represent the civilian, non-institutionalized population of the US. Oral health data are currently available for 8082 and 9010 persons aged \geq 2 years who participated in the 1999–2000 and 2001–2002 NHANES, respectively. This article provides background information on previous national examination surveys with oral health content. It also provides general analytical considerations, oral health content information, and evaluations of data quality in terms of examiner reliability statistics (percent agreements, kappa, and correlation coefficients) for the 1999–2002 NHANES Oral Health Component.

Since 1959, a series of health examination surveys of the US population have been conducted by the National Center for Health Statistics (NCHS), Centers for Disease Control and Prevention (CDC). The initial surveys were named the National Health Examination Surveys (NHES) and focused on adults (age 18–79 years) in 1959– 1962, on children (age 6–11 years) in 1963–1965, and on adolescents (12–17 years) in 1966–1970 (1–4). When an expanded dietary and nutritional assessment was added to the 1970–1972 examination cycle, the name of the survey was changed to the National Health and Nutrition Examination Survey (NHANES) (5). Subsequently, a second

B. A. Dye^{1,2}, L. K. Barker³, R. H. Selwitz⁴, B. G. Lewis¹, T. Wu⁴, C. D. Fryar¹, Y. Ostchega¹, E. D. Beltran³ and E. Ley^{5,6}

¹National Center for Health Statistics, Centers for Disease Control and Prevention, Hyattsville, MD, ²School of Dentistry, University of Maryland, Baltimore, MD, ³Division of Oral Health, National Center for Chronic Disease Prevention and Health Promotion, Centers for Disease Control and Prevention, Atlanta, GA, ⁴National Institute of Dental and Craniofacial Research, National Institutes of Health, Bethesda, MD, ⁵University of Minnesota, Duluth, MN, ⁶Westat, Rockville, MD, USA

Key words: data reliability; dental public health; epidemiology; NHANES; oral health; quality assurance

Bruce Dye DDS, MPH, CDC/NCHS/NHANES Program, 3311 Toledo Road, RM 4416, Hyattsville, MD 20782 Tel: 301 458 4199 e-mail: bfd1@cdc.gov

Submitted 10 November 2005; accepted 7 March 2006

National Health and Nutrition Examination Survey (NHANES II) was conducted in 1976–1980 and NHANES III followed during 1988–1994 (6, 7). During the period 1982–1984, a survey focusing on Hispanic populations in the US was conducted (HHANES) (8). Oral health information has been collected during all these surveys, except for NHANES II.

Beginning in 1999, NHANES was changed from a periodic survey to a continuous, annual survey. Unlike previous surveys, the current NHANES is a nationally representative sample for each year of data collection but data are released in 2-year periods to protect confidentiality and increase statistical reliability. As in previous national health examination surveys, participants are interviewed in their homes and then complete a health examination, which includes the collection of biologic specimens for laboratory testing at a Mobile Examination Center (MEC). Each MEC consists of four interconnected specially designed trailers. Two health examination teams are deployed simultaneously in the field. Each health examination team is comprised of 16 specially trained health professionals and support staff, including a licensed dentist. Many members of the health team are trained to perform multiple tasks, including recording of dental data as called by the dentist. The interview teams, comprised of highly trained home interviewers, conduct the household interviews.

Some of the primary objectives of NHANES include: monitoring trends in the prevalence, awareness, treatment, and control of selected diseases and risk factors in the US; assessing risk behaviors and environmental exposures; and studying the relationship between diet, nutrition, and health. Because the current NHANES is a continuous survey, examination components may be phased in or out as the survey evolves to fulfill existing objectives and to answer emerging public health questions. Excluding laboratory assessments, there are typically more than a dozen examination components requiring over 3 h of examination time for a middle-aged adult to complete. However, examination component eligibility is generally age-based and examination profiles vary among the study participants. All participants aged ≥ 2 years are eligible for an oral health examination. Additional details on the content of examination components can be found elsewhere (http://www.cdc.gov/nchs/about/ major/nhanes/datalink.htm).

Survey design considerations

The target population for the current NHANES is the civilian, noninstitutionalized population of all ages residing in the 50 states and the District of Columbia in USA. As part of the sample design, the NHANES over-samples some population subgroups so that reliable estimates are produced for these groups. Over-sampling for the current NHA-NES includes adolescents 12–19 years, persons \geq 60 years, African-Americans, Mexican-Americans, and persons of the low-income group. The NHANES uses a stratified, multistage probability sampling design to select study participants. Primary sampling units (PSUs), which are usually individual counties, are identified first. Occasionally, contiguous counties are selected to insure that PSUs remain a minimum sample size. The second stage of sample selection is identifying segments, which are a combination of Census blocks. Households are selected from segments and then one or more participants are selected from the households during the fourth stage of sample selection. Study participants are compensated for their time (\$70.00–100.00 for adults and \$35.00–40.00 for children) upon completion of their examination.

Trained NHANES interviewers administer screening questionnaires and select one or more participants from eligible households. Once eligible participants are identified by the interviewer, recruitment begins with assurance of data confidentiality and informed consent procedures. The household interview has two components: the household and sample person interviews. Following completion of the household interview, the interviewer reviews the health examination component with the participant(s). All interviewed participants are asked to take part in the health examination in a MEC and undergo a second series of informed consent procedures for the health examination.

Sample design characteristics for previous and current NHANES are shown in Table 1. The current NHANES sampling domains are defined by age, sex, and race/ethnicity. The age categories for race/ethnicity are 0–11 months, 1–2, 3–5, 6–11, 12–15, 16–19, 20–39, 40–59 and \geq 60 years. The number of participants who completed various portions of the current NHANES is shown in Table 2. Among persons aged \geq 2 years who completed the home interview, approximately 88% completed the oral health examination in the 1999–2000 NHANES. Similarly, 89% completed the dental examination in the 2001–2002 NHANES.

Statistical considerations

Because NHANES uses a complex survey design involving unequal probabilities for selection, complex estimation procedures are required to produce valid population estimates. In the final dataset, weights account for the unequal probability

Table 1. Comparison of sampling	g design characteristics for pre	vious and current National Healt	th and Nutrition Examine	ation Surveys (NHANES) w	ith oral health content
Characteristic	NHANES I	Hispanic HANES	NHANES III	NHANES 1999–2000	NHANES 2001-2002
Data collection year Age of the target population Dental examination age Number of survey examination Locations	1971–1974 1–74 years 1–74 years 65	1982–1984 6 months to 74 years 2–74 years 30	1988–1994 2 months and older 1 year and older 89	1999–2000 From birth 2 years or older 27	2001–2002 From birth 2 years or older 30
Eligible geographical areas for sample	48 lower continental states	Southwest for Mexican Americans; NY, NJ, CT for Puerto Ricans; Dade county Florida for Cubans	50 states	50 states + DC	50 states + DC
Groups targeted for expanded sampling	Persons 65 years or older; children 1–5 years of age; women aged 20–44 years	Persons 6 months to 19 years; and those aged 45–74 years	Persons 2 months <6 years; non-Hispanic blacks; Mexican-Americans; those over 60 years	Persons 12–19 years; African-Americans; Mexican-Americans; those 60 years or older; and others ^{a, b}	Persons 12–19 years; African-Americans; Mexican-Americans; those 60 years or older; low-income whites: and others ^b
Total interviewed Total MEC examined ^c Total with a dental examination record	28 043 20 749 20 218	13 689 11 653 11 347	33 994 30 818 28 059	9965 9282 8082	11 039 10 477 9010
^a Low-income whites were over s: ^b All women who indicated that ti ^c Mobile Examination Center (ME	ampled only for NHANES 20 hey were pregnant at time of C).	00. screening were included in the s	sample.		

1999-2002 by selected de	emograph	ic char	acterist	ics						
	1999/200	00				2001/200)2			
	Number	sampl	e perso	ns ^a	Evamination	Number	sample	persons	^a	Examination
Characteristic	Selected	HIQ	MEC ^c	OHX	weight (mean)	Selected	HIQ	MEC ^c	OHX	weight (mean)
Age										
2–5 years	954	822	768	690	21 213	1075	987	951	899	16 327
6–11 years	1262	1113	1061	1024	23 448	1404	1242	1202	1174	20 560
12–19 years	2732	2415	2314	2220	13 679	2799	2487	2418	2280	13 377
20–39 years	2127	1695	1569	1469	53 835	2353	1925	1843	1693	44 384
40–59 years	1791	1351	1269	1188	51 839	2065	1614	1556	1444	49 528
60+ years	2483	1834	1606	1491	25 999	2493	1872	1628	1520	25 019
Sex										
Male	5559	4480	4185	3957	30 803	5894	4870	4650	4389	28 404
Female	5790	4750	4402	4125	30 875	6291	5257	4948	4621	28 316
Race/ethnicity ^b										
Non-Hispanic Blacks	2599	2107	2001	1878	15 536	2843	2494	2398	2214	9616
Mexican-Americans	3484	3066	2908	2704	6957	2904	2468	2387	2241	13 535
Non-Hispanic Whites	4203	3181	2877	2740	61 910	5428	4297	4005	3792	46 397
Others	1063	876	801	760	44 182	1014	868	808	763	38 362
Total	11 349	9230	8587	8082	_	12 189	10 127	9598	9010	-

Table 2. Selected examination weight values and the number of sampled persons aged 2 years or older completing Interview, MEC and Oral Health Examinations for the National Health and Nutrition Examination Survey (NHANES) 1999–2002 by selected demographic characteristics

^aNumber of sampled persons Selected to participate in NHANES; completing a Home Interview Questionnaire (HIQ); completing a Mobile Examination Center examination (MEC); and completing an Oral Health examination (OHX). ^bRIDRETH1 race/ethnicity variable.

^cMobile Examination Center (MEC).

-, Not reported.

of selection and nonresponse, and include poststratification to US Census estimates of the general population. The NHANES public-use dataset includes individual sample weights for the Interview and MEC Examination for 2-year data analyses using either 1999–2000 or 2001–2002 data as well as for an aggregated 4-year data analysis. The number of participants who completed the home interview and MEC examination, including the oral health component as well as mean values for the examination weights for selected demographic characteristics is presented in Table 2.

An earlier public release of the 1999–2000 NHA-NES data did not include information on PSU and stratification because of confidentiality issues; thus, the 1999–2000 dataset initially included 52 replicate weights to be used in a jackknife technique to estimate variance (9). Subsequently, masked variance units (MVUs) were derived from the PSU information to minimize confidentiality concerns and have been made available for the 1999–2000 dataset. These MVU variables have also been made available for the 2001–2002 cycle and in the combined 1999–2002 dataset to allow for variance estimation using Taylor series linearization techniques. Software such as SUDAAN or STATA can be used to estimate sampling errors by the Taylor series method. Additional analytical guidelines for the current NHANES are available elsewhere (http:// www.cdc.gov/nchs/data/nhanes/nhanes_03_04/ nhanes_analytic_guidelines_dec_2005.pdf). An additional resource for analysts is the Dental, Oral and Craniofacial Data Resource Center (DRC) (http:// drc.hhs.gov/). Programming code and guidance will be made available to promote standardized use of derived variables as well as future publication initiatives and announcements.

Investigators seeking to conduct analytical work or multistrata descriptive analysis should carefully monitor the number of observations in the subcategories and the effect of influential observations when using 2-year datasets. In many cases, it may be optimal to use ≥ 4 years of data for analysis. Moreover, there are many issues relevant to reporting estimates derived from small sample sizes. A general guideline for determining when NHANES data may be considered statistically unreliable is when an estimate is based on fewer than 30 sample units in the denominator or if the relative standard error is >30%. Unreliable estimates should either not be reported, or at a minimum, identified as unreliable.

Oral health component

For the current NHANES, oral health information was collected during the home interview and the physical examination administered in the MEC. The specific objectives and related data uses for the NHANES oral health data include:

- Assessing the prevalence of major oral health diseases and conditions including dental caries, periodontal disease, dental trauma, dry mouth, and dental fluorosis.
- Assessing prevention and treatment efforts including the prevalence of dental sealants.
- Monitoring the oral health status of minority/ underserved populations.
- Evaluating *Healthy People 2000 and 2010* objectives related to oral health.
- Supporting research activities as identified in the 2000 Surgeon General's Report on *Oral Health in America*.

The overall target population for the oral health component of the current NHANES was persons ≥2 years of age. Participation in the various subcomponents was determined by age, which was related to the disease or condition being examined or queried. Table 3 shows the age eligibility requirements and the sequence of administration for the various oral health examination subcomponents and home interview for the 1999–2002 NHANES. Additional oral health subcomponent modifications have been made to the 2003–2004 NHANES, which will be described in a future publication.

Before a trained dentist performed the examination, a medical history screening was conducted by the dentist to identify participants who were to be excluded from the periodontal and root caries assessments. A screening questionnaire was administered by proxy for persons ≤15 years of age. A positive response to any of the health screening questions would indicate exclusion (see Appendix). Approximately 9.0% of the dental examination participants were excluded because of medical history concerns from portions of the oral health examination in 1999-2000 and 6.9% were excluded in 2001–2002. Among those excluded, approximately 33% and 38% during 1999-2000 and 2001-2002, respectively, resulted from a positive response to the question inquiring the need for antibiotic prophylaxis.

Following the completion of the medical history screening, the oral health examination began. An orofacial pain assessment was initiated by the Table 3. Oral health subcomponents age eligibility and sequence of administration, 1999–2002 NHANES

Assessment	1999–2000 (age in years)	2001–2002 (age in years)
Examination	5	<u> </u>
Medical history	12+	13+
Orofacial pain	10-69	10-69
Tooth count	2+	2+
Coronal caries	2+	2+
Root caries	18+	18+
Dental sealants	2–34	2–34
Dental fluorosis	6–49	6–49
Incisor trauma	10-29	10–29
Gingival assessment	12–49	_
Periodontal assessment ^a	18+	13+
Bleeding from probing ^b	_	13+
Recommendation for care	2+	2+
Interview		
Dental health		
perception, dental visits		
And dental care utilization	2+	2+
Dry mouth and chewing problems	40+	18+

-, An assessment was not performed during the period. ^a1999–2000 periodontal assessment: mesial and midfacial sites; 2001–2002 assessments: mesial, mid-facial, and distal sites.

^bBleeding from probing was ascertained from all three probed sites.

dental recorder with a brief questionnaire inquiring about pain, sensations, and irritations in the orofacial area as well as the frequency and quality of life impact of the pain experienced. A measurement of maximal incisal opening was made by the dentist and the temporomandibular joint (TMJ) area was assessed by palpating the masseter and preauricular muscles for tenderness. The assessment for surgical dental implants as well as the presence or absence of permanent and/or primary teeth for each of the 32 tooth positions were performed during the tooth count. Coronal and root caries were assessed using the visual-tactile method. The presence of dental restorations or replacements for missing permanent teeth was also ascertained. To assign appropriate codes for tooth loss and replacement, the dentist asked participants about the reasons for missing permanent teeth. Assessments of root caries and restorations indicated whether one or more lesions (restorations) were present in the mouth. The first and second primary and permanent molars, the premolars and upper lateral permanent incisors were assessed for dental sealants. The presence of incisor trauma required a

history of trauma and clinical signs of trauma or repair. The dental fluorosis assessment was performed on all teeth, excluding the third molars, using a modified Dean's Fluorosis Index method (http://www.cdc.gov/nchs/data/nhanes/ oh-e.pdf).

A periodontal examination was performed on two randomly selected quadrants (one maxillary and one mandibular) using a color-banded NIDR probe graduated at 2, 4, 6, 8, 10, and 12 mm. The first measurement was the distance between the free gingival margin (FGM) and the cementoenamel junction (CEJ). The second measurement was from the FGM to the bottom of the pocket (PD). Measurements were made at the mesio-facial and mid-facial sites, and a disto-facial site was added in 2001. The oral health examination concluded with the recommendation for care. The dentist assigned each participant to one of four care levels: see a dentist immediately, see a dentist within 2 weeks, see a dentist at the earliest convenience, and continue routine dental care. The dentist provided an overview of her/his findings directly to the participant and a report of findings was provided to them during the MEC exiting procedures. If an SP required urgent dental attention, contact information for local public health and community clinics were provided to the participant if (s)he reported not having a regular source of dental care, such as a family dentist, a clinic, etc. Details of the examination protocols and procedures for the 1999-2000 and 2001-2002 NHANES oral health component are available elsewhere (http://www.cdc.gov/nchs/ data/nhanes/oh-e.pdf).

Dental team, training, and field testing

Dental examiners for the 1999–2002 NHANES were dentists licensed in at least one US jurisdiction. Other MEC personnel were trained as dental recorders. The current NHANES was field-tested in 1998, and initial training and calibration of the dental examiners was provided under the leadership of Dr Ley (coauthor), who was the reference examiner during the entire NHANES III period. In 2000, a new reference examiner (author) was trained by Dr Ley for the current NHANES.

The majority of all dental examinations (91%) were performed by dental examiners assigned to the MEC examination teams. However, to minim-

ize the loss of oral health data when the primary dental examiner was unavailable, back-up dental examiners were periodically deployed to provide coverage during examiner absences of >24 h. All back-up examiners received the same rigorous training and calibration prior to functioning independently during survey data collection. The reference examiner also provided extended coverage as needed. Approximately 9% of dental examinations were completed by the back-up dental examination team and the reference examiner from 1999 to 2002. The composition of the back-up pool of dental examiners over the current survey period included five US Public Health Service dental officers and one dentist, subcontracted through the NHANES prime contractor (Westat), who had served previously as a primary dental examiner for NHANES III.

Modifications and technical notes

The oral health examination was modified in 2001 and again in 2002. Beginning in 2001, the gingival sweep used to assess gingival bleeding was replaced with assessments for bleeding from probing. Additionally, a third probing site (distofacial) was added and the age of eligibility for participation in the periodontal assessment was changed from 18 years to \geq 13 years. There were also slight location changes for some of the other probing sites. The mid-facial probe site for posterior molars was moved to correspond directly to the furcation area and the mesio-facial probe site was moved slightly more interproximally. Beginning in 2002, there were three additions to the dentition assessment: (i) the presence of residual dental roottips was assessed during tooth count; (ii) the type of replacement for missing teeth (i.e. either removable or fixed) was ascertained in coronal caries; and (iii) denture questions were added. These items were not released in the 2001-2002 NHANES oral health dataset because they were added during the middle portion of the data collection cycle. Consequently, derived variables were created for data collected during the 2001-2002 NHANES to make some of the 2002 variables comparable with the 2001 variables. This was done for tooth count, coronal tooth condition, and coronal surface condition. Residual dental root tips were recoded as permanent teeth present with all surfaces decayed and missing teeth replaced with either a fixed or a removable prosthesis were recoded as simply

'replaced.' Data from the denture questions were not released.

In 2002, the medical exclusion question inquiring into a history of need for taking antibiotics prior to dental treatment was modified slightly. With the beginning of the current NHANES (1999), all dental examiners were instructed to verbally probe further if a positive response was given to this question. All probes were to include the word 'always' needed to take antibiotics prior to dental treatment or dental care. Consequently, 'always' was added to the question's text to reflect accurately what the respondents were reporting 'yes' to.

Investigators planning to analyze NHANES 1999-2002 data should be familiar with the structure of the dataset, including the variables, codes, and the timeframes of the modifications. The datasets are organized into 'chapters,' which are labeled as 'Dentition', 'Periodontal', 'Miscellaneous', and 'Home Interview' files. For the 2001–2002 NHANES oral health dataset, periodontal information is released in two separate chapters labeled as 'Periodontal-Maxillary' and 'Periodontal-Mandibular' files. The 1999-2000 and 2001-2002 oral health datasets are available at: http://www.cdc.gov/ nchs/about/major/nhanes/datalink.htm. To assist data users and promote the use of commonly derived variables for analysis, recommended derivations and SAS sample code are available from the NIDCR/CDC Dental, Oral, and Craniofacial Data Resource Center at: http://drc.hhs.gov/.

Quality assurance

The current NHANES used automated data collection procedures with all data recorded directly onto a computerized data collection form to reduce error. The quality of oral health data in this survey was controlled by an intense training period for dental examiners and recorders, calibration of dental examiners, and periodic monitoring and recalibration of dental examiners. In the training phase, the reference examiner (trainer) used lecture and slide presentations on each assessment to familiarize examiners with the study protocols and research criteria, including data recording and editing for that assessment. Infection control and emergency preparedness guidelines were reviewed as well. Demonstrations of examination technique and equipment use were also conducted.

In the standardization phase, the reference examiner and examiners-in-training examined the same set of volunteers. During this period, trainees were encouraged to ask questions regarding criteria while conducting the study protocols. A detailed discussion of observations was led by the reference examiner following each standardization round of examinations with the intent to systematically minimize differences in examination findings. The final phase of training included a preliminary calibration cycle with a follow-up calibration conducted in the field shortly after the dental examiner arrived for his/her first MEC stand. During calibration, examiners performed independent replicate examinations without discussion. Data from the calibration sessions were analyzed to measure consistency between each examiner and the reference examiner. Training typically lasted for 40 h and was conducted in the Washington, DC, metropolitan area. The follow-up field calibration cycle was performed during a MEC's normal examination schedule and typically lasted for five to six 4-h examination sessions (20-24 h), depending on the number of study participants scheduled for those MEC sessions.

The reference examiner visited each dental examiner two to three times each year to observe field operations and to replicate randomly 20-25 dental examinations during each visit. The purpose of these periodic visits was to determine if the examiners were conducting the oral health examinations within the parameters of the study protocols and if the standard for examination between the examiner and reference examiner had been maintained. Data from these replicate examinations were used to produce inter-rater reliability statistics. If an examiner's performance fell below an acceptable level, retraining was conducted on site. An examiner's performance was also assessed by monitoring results of second examinations. Approximately 5% of eligible study participants were asked to return to the MEC at a later time to have a number of assessments repeated, including the oral health exam. Data from the second exams were compared to the primary examination data to produce intra-rater reliability statistics.

For this report, weighted and non-weighted kappa statistics were produced using SAS software (Version 8.02; SAS Institute Inc, Cary, NC, USA). *Healthy People 2010*-derived oral health variables were used to produce inter- and intrarater reliability statistics for untreated primary and permanent decay, primary and permanent caries experience, tooth retention, dental sealants, and periodontal disease. A Dean's Fluorosis Index derived variable was used for comparisons. For incisor trauma, individual kappa statistics were produced for each tooth position and a subjectlevel mean score was calculated. Inter- and intraclass correlation coefficients (ICCs) were calculated using subject-level mean values (mm) for loss of attachment and pocket depth. Measurements obtained from the mesial and mid-facial periodontal sites were combined to produce subjectlevel means for both data collection cycles (1999-2000 and 2001-2002). Additional reliability statistics for the oral health component are available from the authors.

Table 4 shows selected inter- and intra-rater reliability statistics for the 1999-2002 NHANES oral health examination. The reliability statistics are presented for the two primary dental examiners employed during the 1999-2000 data collection cycle and for the three primary dental examiners employed during the 2001-2002 period. Examiner B is the same person in both periods (1999-2000 and 2001–2002), whereas examiner A represented a different examiner for each of the two data collection periods and examiner C only participated in the latter data collection cycle. The inter-rater reliability values expressed as kappa scores varied from 0.48 to 1.00 for dental caries in permanent and primary teeth. The intra-rater reliability values varied from 0.70 to 1.00 during 1999-2002 for caries in permanent teeth. The inter- and intra-rater reliability values for dental fluorosis ranged from 0.56 to 0.73 during 1999-2002. The inter-rater strength of agreement observed for dental sealants varied from 0.72 to 0.93. For the *Healthy People 2010* variable for prevalence of periodontal disease in the US (loss of attachment >4 mm at one or more sites), the strength of agreement between the primary dental examiners and the reference examiner from 1999-2002 varied from 0.58 to 0.77.

For the overall subject-level inter- and intra-class correlation coefficients (ICC) measures, the interrater reliability ranged from 0.724 to 0.893 for loss of attachment and 0.550–0.866 for pocket depth (Table 5). Loss of attachment intra-class coefficients varied from 0.876 to 0.973. An analysis of potential examiner bias (systematic differences) regarding mean differences in loss of attachment/pocket depth between examiner and the reference examiner was performed. Mean differences in measurements across all periodontal sites ranged from 0.015 to 0.286 mm for attachment loss and from 0.013 to 0.524 mm for pocket depth.

Discussion

Percent agreements, kappa scores and ICCs are measures of examiner reliability. Different researchers have assigned various standards for the interpretation of these values. For example, the most widely known guideline for evaluating the strength of agreement for kappa scores is that proposed by Landis and Koch: ≤0 is 'poor', 0.01-0.20 is 'slight', 0.21-0.40 is 'fair', 0.41-0.60 is 'moderate', 0.61-0.80 is 'substantial', and >0.80 is 'almost perfect' (10). Although other interpretations of kappa exist, all are basically subjective (11). Nearly two decades ago, Hunt reported that 'standards for acceptable reliability on dental measures had not been developed' and suggested a three-tier range for kappa score interpretation: below 0.40 is 'poor' agreement, 0.40-0.75 represents 'fair-good' agreement, and >0.75 represents 'excellent' agreement (12). Similar interpretations for ICCs also exist and the ICC has been shown to be comparable with the kappa statistic (13).

Although a variety of reliability statistics are reported in this paper reflecting different types of data collected or derived (nominal, continuous, etc.) for the 1999–2002 NHANES, the interpretation of these reliability statistics should depend upon the context of the study design, data use, and clinical relevance. Important determinates for the magnitude of reliability statistics include prevalence for kappa (11) and if observations vary little in score, the ICC tends to be low because of the comparison made between variance among observations to the total variance (14).

There have been previous reports describing some measures of quality for a variety of oral health indicators regarding the NHANES III data (15–19). As a result of comparing prevalence estimates between phase 1 and phase 2 of NHA-NES III, one of these reports expressed concerns that a portion of the NHANES III periodontal data was biased (17). In response to the data quality issues raised, Winn et al. (18) outlined a number of survey design factors, including sampling variation, that were more likely to influence the observed prevalence differences between the two NHANES III phases. Moreover, the authors cautioned that differences in some results by survey

Table 4. Summ	ary den	tal examine	er inter-rate	er ^a and in	tra-rater	^b reliability	statistics fo	or selected	d denta	al characteris	stics: NHA	NES, 1999	9–2002			
	1999–2	2000 Inter-re	ater		1999–2	2000 Intra-rat	er		2001-	-2002 Inter-ra	ter		2001–2	.002 Intra-rat	er	
Characteristic	n^{g}	% Agree	Kappa	$\mathrm{ASE}^{\mathrm{h}}$	n^g	% Agree	Kappa	ASE^{h}	n^g	% Agree	Kappa	$ASE^{\rm h}$	n^{g}	% Agree	Kappa	$\mathrm{ASE}^{\mathrm{h}}$
Untreated prime Examiner A Examiner B	ry decay 27 10	رد 100 100	$1.00 \\ 1.00$	0.00	4 4	1 1	1.1	11	15 17 8	100 82.4 100	1.00 0.48 1.00	0.00 0.23 0.00	10%	1 1 1	1 1 1	1 1 1
Primary caries e Examiner A Examiner B Examiner C	xperienc 27 10	e ^c 100 100	$1.00 \\ 1.00$	0.00	44	1 1	1 1	1 1	115 17 8	$100 \\ 88.2 \\ 100 \\ 100 $	1.00 0.75 1.00	0.00 0.16 0.00	0 1 0 0			
Untreated permulticated permulticated bermuiter A Examiner B Examiner C	anent de 100 63	cay ^c 98.0 95.2	$0.94 \\ 0.89$	0.04 0.06	$\frac{114}{138}$	91.2 94.2	0.79 0.86	0.06	97 58 55	92.8 93.1 94.5	$\begin{array}{c} 0.73 \\ 0.70 \\ 0.84 \end{array}$	$\begin{array}{c} 0.10 \\ 0.14 \\ 0.09 \end{array}$	$\begin{array}{c} 176\\74\\116\end{array}$	96.0 90.5 92.2	$\begin{array}{c} 0.89 \\ 0.64 \\ 0.71 \end{array}$	$\begin{array}{c} 0.04 \\ 0.12 \\ 0.09 \end{array}$
Permanent carie Examiner A Examiner B Examiner C	s expent 100 63 r - d.	ence 98.0 95.2	0.96 0.90	0.03 0.06	114 138	96.5 97.1	06.0 0.90	0.05 0.05	97 58 55	97.9 93.1 96.4	0.96 0.85 0.93	0.03 0.07 0.05	176 74 116	97.7 93.2 99.1	0.92 0.84 0.97	$\begin{array}{c} 0.04 \\ 0.07 \\ 0.03 \end{array}$
Dean's Fluorosis Examiner A Examiner B Examiner C	1ndex 84 57	70.2 68.4	0.73 0.70	0.05 0.06	87 92	64.4 85.9	0.57 0.72	0.07 0.07	79 46 44	65.8 67.3 70.5	$\begin{array}{c} 0.60 \\ 0.67 \\ 0.61 \end{array}$	09.0 0.07 0.09	132 51 91	63.6 88.2 78.0	0.56 0.72 0.65	$\begin{array}{c} 0.07 \\ 0.07 \\ 0.07 \end{array}$
Examiner A Examiner A Examiner B Examiner C	76 50	98.7 97.2	0.98 0.96	0.02 0.05	61 67	97.7 93.6	0.97 0.91	0.04 0.07	72 42 41	99.0 99.1 100	$\begin{array}{c} 0.95 \\ 0.99 \\ 1.00 \end{array}$	0.05 0.02 0.00	76 34 60	98.0 99.6 100	0.96 0.98 1.00	$\begin{array}{c} 0.04 \\ 0.01 \\ 0.00 \end{array}$
Examiner A Examiner A Examiner B Examiner C	ce ⁻ 66 46	97.0 93.5	0.93 0.82	0.05 0.10	56 69	92.9 97.1	0.67 0.86	$0.15 \\ 0.10$	63 34 37	93.7 91.2 94.6	$\begin{array}{c} 0.83 \\ 0.72 \\ 0.85 \end{array}$	$\begin{array}{c} 0.08 \\ 0.15 \\ 0.10 \end{array}$	85 31 61	97.7 83.9 96.7	0.90 0.51 0.89	$\begin{array}{c} 0.07 \\ 0.19 \\ 0.08 \end{array}$
Examiner A Examiner A Examiner B Examiner C	ase 51 34	88.2 91.2	0.71 0.68	$0.11 \\ 0.17$	81 104	88.9 94.2	0.66 0.78	0.11 0.09	78 46 41	91.0 89.1 95.1	$\begin{array}{c} 0.58 \\ 0.67 \\ 0.77 \end{array}$	$\begin{array}{c} 0.14 \\ 0.13 \\ 0.15 \end{array}$	$\begin{array}{c} 161 \\ 75 \\ 108 \end{array}$	91.9 89.3 97.2	$\begin{array}{c} 0.65 \\ 0.74 \\ 0.81 \end{array}$	$\begin{array}{c} 0.09\\ 0.09\\ 0.11\end{array}$
Looun retention Examiner A Examiner B Examiner C	100 63	97.0 96.8	0.90 0.88	0.05 0.08	114 139	95.6 94.2	0.90 0.88	$0.04 \\ 0.04$	97 58 55	100 93.1 98.2	$1.00 \\ 0.64 \\ 0.95$	$\begin{array}{c} 0.00\\ 0.16\\ 0.05\end{array}$	176 74 116	98.9 98.6 96.6	0.97 0.97 0.91	$\begin{array}{c} 0.02 \\ 0.03 \\ 0.04 \end{array}$
^a Primary examir ^b First and secon	her and r I examir	eference ex vation comp	aminer con varisons.	iparisons.								•	•			

^cHP2010 Objective: For untreated decay and caries experience, proportion of people with at least one tooth affected by the condition. For dental sealants, proportion of people with at least one dental sealant present on a their permanent molar teeth (excluding third molars). For periodontal disease, proportion of people with attachment loss of 4 mm or more at one or more sites measured. ^dRecommended derived variable – code available in 2001–2002 data release notes. ^eWeighted kappa. ^fPooled kappa. ^fAsymptotic standard error. ^bAsymptotic standard error. ^bAsymptotic standard error.

Table 5. Dentai	exan	niner inter-clé	ass ^a and intra	t-class ^b c	orrelat	tion coeffici	ients (ICC) for	r overall	mean	I loss of attac	hment and n	nean po	cket de	epth: NHA	NES, 1999–2	002
	1999	9−2000 Inter-1	rater		1999-	-2000 Intra-	rater		2001	-2002 Inter-ra	ater		2001-	-2002 Intra-	-rater	
Characteristic	^p u	Reference mean	Examiner mean	ICC	p ^q	First Ex mean	Second Ex mean	ICC	^p u	Reference mean	Examiner mean	ICC	^p u	First Ex Mean	Second Ex Mean	ICC
LOA mean ^c																
Examiner A	51	1.168	0.909	0.846	81	0.609	0.591	0.918	78	0.375	0.360	0.885	161	0.567	0.532	0.876
Examiner B	34	0.766	0.620	0.725	104	0.732	0.763	0.884	46	0.697	0.833	0.893	75	0.870	0.868	0.940
Examiner C									41	0.399	0.219	0.724	108	0.300	0.282	0.973
PD mean	ì				5	1 100					100		777			
Examiner A	10	1.626	1.613	0.866	81	1.498	1.54/	0.749	8	C8/.0	0.811	0.744	161	1.063	1.035	0.827
Examiner B	34	1.256	1.009	0.550	104	1.158	1.178	0.756	46	1.072	1.337	0.600	75	1.241	1.256	0.865
Examiner C									41	0.884	0.656	0.637	108	0.700	0.680	0.827
^a Primary exami	iner ar	nd reference (examiner con	nparison	s.											

²First and second examination comparisons. ²First and second examination comparisons. ⁶Overall mean is for the combined mesial and mid-facial periodontal sites. ^dNumber of paired observations. years in a continuous survey could be a significant issue as well. Nevertheless, it is important to remember that sampling and measurement variations will occur in any survey.

Changing NHANES to a continuous survey with constant content for 2 years allows for more flexibility in data collection while still tracking key oral health indicators and conditions. Oral health assessments can be added and withdrawn periodically to maximize needed sample size to produce reliable estimates. For instance, oral conditions with a higher prevalence may need only 2 or 4 years of data collection to achieve a desirable sample size, whereas oral conditions that are rarer may need ≥ 6 years of continuous data collection. Moreover, changing the mix of oral health assessments more frequently will allow for additional opportunities to explore interrelationships between a variety of oral conditions.

This paper describes a broad range of data quality issues including values for examiner concordance for the 1999-2002 NHANES oral health component. Five primary dental examiners collected most of the 1999-2002 NHANES oral health data compared with three examiners for NHANES III. Although the attrition of highly trained dental examiners was higher for the current NHANES compared with NHANES III, the disruption to data collection was minimized: (i) by drawing from a pool of previously trained back-up examiners and (ii) by relying on existing training systems established to operate in a continual data collection environment. Another contributory factor toward enhancing data quality of the current NHANES that was not employed on NHANES III was direct data entry and operating ongoing quality assurance programs during data collection instead of relying on postdata collection quality assurance activities alone.

Most of the current NHANES oral health examination methods were structured to be consistent with previous dental examination procedures used in other national surveys conducted by NCHS and NIDCR to allow for the monitoring of trends in oral health status across many surveys. Moreover, the oral health datasets are linked to a multitude of other current NHANES data components covering a broad range of health, behavior, and physical status. Because of this linkage, the current NHANES oral health data can be used to investigate a plethora of systemic interrelationships which would often be too resource-intensive to be conducted in smaller, independent clinical studies.

Acknowledgements

The 1999–2002 NHANES oral health component was a funding and content collaborative effort between the NIH/National Institute of Dental and Craniofacial Research, the CDC/National Center for Health Statistics, and the CDC/National Center for Health Promotion and Disease Prevention Division of Oral Health. The authors also acknowledge Beth Hintz, the Westat NHANES Oral Health Project Officer, for reviewing and providing comments to this manuscript.

At the time this report was prepared, Dr Robert Selwitz was employed with the NIH/National Institute of Dental and Craniofacial Research, Bethesda, MD. He is currently employed with University of Florida College of Dentistry and the Duval Country Health Department, Jacksonville, FL.

Ethics Statement

All study participants gave informed consent in accordance with the Institutional 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.

References

- 1. US Department of Health, Education, and Welfare, National Center for Health Statistics. Vital health statistics. Series 1, Number 1. Origin, program, and operation of the US National Health Survey. Washington, DC: National Center for Health Statistics; 1965.
- 2. US Department of Health, Education, and Welfare, National Center for Health Statistics. Vital health statistics. Series 1, Number 4. Plan and initial program of the Health Examination Survey. Washington, DC: National Center for Health Statistics; 1965.
- 3. US Department of Health, Education, and Welfare, National Center for Health Statistics. Vital health statistics. Series 1, Number 5. Plan, operation, and response results of a program of children's examinations. Washington, DC: National Center for Health Statistics; 1967.
- 4. US Department of Health, Education, and Welfare, National Center for Health Statistics. Vital health statistics. Series 1, Number 8. Plan and operation of a Health Examination Survey of US youths 12–17 years of age. Washington, DC: National Center for Health Statistics; 1974.
- 5. US Department of Health, Education, and Welfare, National Center for Health Statistics. Vital health statistics. Series 1, Number 10. Plan and operation of the Health and Nutrition Examination Survey, US 1971–73. Rockville, MD: National Center for Health Statistics; 1973.
- 6. US Department of Health and Human Services, National Center for Health Statistics. Vital health statistics. Series 1, Number 15. Plan and operation of the second National Health and Nutrition Examina-

tion Survey, 1976–80. Hyattsville, MD: National Center for Health Statistics; 1981.

- 7. US Department of Health and Human Services, Centers for Disease Control and Prevention, National Center for Health Statistics. Vital health statistics. Series 1, Number 32. Plan and operation of the third National Health and Nutrition Examination Survey, 1988–94. Hyattsville, MD: National Center for Health Statistics; 1994.
- 8. US Department of Health and Human Services, Centers for Disease Control and Prevention, National Center for Health Statistics. Vital health statistics. Series 1, Number 19. Plan and operation of the Hispanic Health and Nutrition Examination Survey, 1982–84. Hyattsville, MD: National Center for Health Statistics; 1985.
- 9. Wolter KM. Introduction to variance estimation. New York: Springer-Verlag; 1990.
- 10. Landis JR, Koch GG. The measurement of observer agreement for categorical data. Biometrics 1977;33:159–74.
- 11. Sim J, Wright CC. The Kappa statistic in reliability studies: use, interpretation, and sample size. Phys Ther 2005;85:257–68.
- 12. Hunt RJ. Percent agreement, Pearson's correlation, and kappa as measures of inter-examiner reliability. J Dent Res 1986;65:128–30.
- 13. Fleiss JL, Cohen J. The equivalence of weighted kappa and the intraclass correlation coefficient as measures of reliability. Educ Psychol Meas 1973;33:613–9.
- 14. Deyo RA, Diehr P, Patrick DL. Reproducibility and responsiveness of health status measures: statistics and strategies for evaluation. Controlled Clin Trials 1991;12:1425–158S.
- 15. Kaste LM, Selwitz RH, Oldakowski RJ, Brunelle JA, Winn DM, Brown LJ. Coronal caries in the primary dentition of children and adolescents 1–17 years of age: United States, 1988–91. J Dent Res 1996;75(special issue):631–41.
- 16. Drury TF, Winn DM, Snowden CB, Kingman A, Kleinman DV, Lewis B. An overview of the oral health component of the 1988–91 National Health and Nutrition Examination Survey (NHANES III, phase 1). J Dent Res 1996;75(special issue):620–30.
- 17. Slade GD, Beck JD. Plausibility of periodontal disease estimates from NHANES III. J Public Health Dent 1999;59:67–72.
- 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.
- 19. Kingman A, Albandar JM. Methodological aspects of epidemiological studies of periodontal diseases. Periodontology 2000;29:11–30.

Appendix

Medical History Screening Questions

• Has a doctor or dentist ever told you that you must always take antibiotics (e.g. penicillin) before you get a dental check up or care? I would like to read a list of health conditions that some people have. As I read off each condition, please tell me whether or not a doctor has ever told you that you have the condition. Has a doctor ever told you that you have...

- A heart problem (specifically congenital heart murmurs, heart valve problems, congenital heart disease, or bacterial endocarditis)?
- Do you have rheumatic fever?

- Kidney disease requiring renal dialysis?
- Hemophilia?
- Pacemaker or automatic defibrillator or artificial material in your heart veins or arteries?
- A hipbone or joint replacement?

[Any positive response would result in the participant being excluded from the root caries and periodontal disease assessments] This document is a scanned copy of a printed document. No warranty is given about the accuracy of the copy. Users should refer to the original published version of the material.