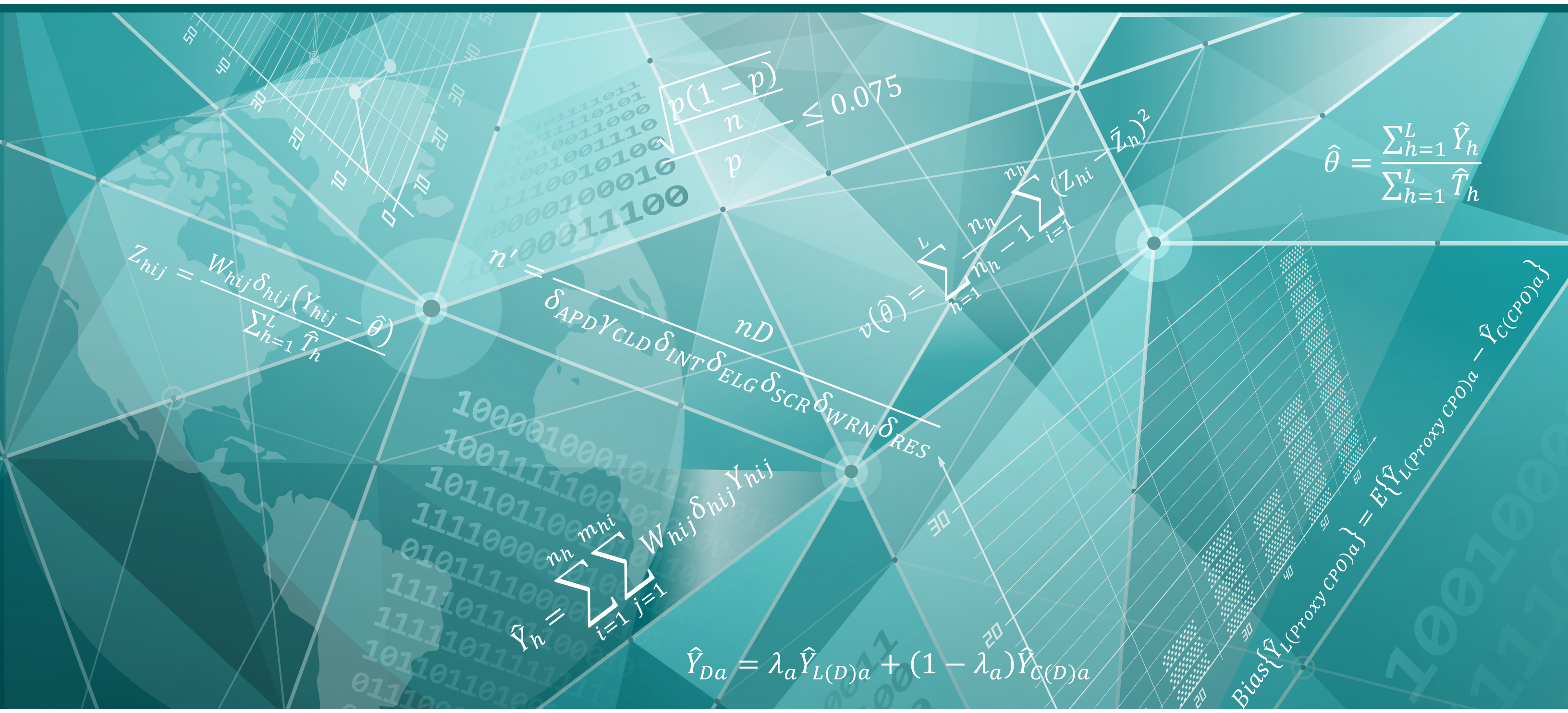
NATIONAL CENTER FOR HEALTH STATISTICS

Vital and Health Statistics

Series 2, Number 183

April 2019



Data Quality Evaluation of the Dental Fluorosis Clinical Assessment Data From the National Health and Nutrition Examination Survey, 1999–2004 and 2011–2016

Data Evaluation and Methods Research

U.S. DEPARTMENT OF HEALTH AND HUMAN SERVICES
Centers for Disease Control and Prevention
National Center for Health Statistics

NCHS reports can be downloaded from: https://www.cdc.gov/nchs/products/index.htm.

Suggested citation

National Center for Health Statistics, National Center for Chronic Disease Prevention and Health Promotion. Data quality evaluation of the dental fluorosis clinical assessment data from the National Health and Nutrition Examination Survey, 1999–2004 and 2011–2016. National Center for Health Statistics. Vital Health Stat 2(183). 2019.

Library of Congress Cataloging-in-Publication Data

Names: National Center for Health Statistics (U.S.), issuing body.
Title: Data quality evaluation of the dental fluorosis clinical assessment data from the National Health and Nutrition Examination Survey, 1999-2004 and 2011-2016.
Other titles: Vital and health statistics. Series 2, Data evaluation and methods research ; no. 183. | DHHS publication ; no. 2019-1383.
Description: Hyattsville, Maryland : U.S. Department of Health and Human Services, Centers for Disease Control and Prevention, National Center for Health Statistics, April 2019. | Series: Vital and health statistics. Series 2, Data evaluation and methods research ; number 183 | Series: DHHS pub ; number 2019-1383 | Includes bibliographical references.
Identifiers: LCCN 2019019953 | ISBN 0840606958
Subjects: | MESH: National Health and Nutrition Examination Survey (U.S.) | Fluorosis, Dental | Evaluation Studies as Topic | Dental Health Surveys | United States | Statistics
Classification: LCC RK52.8 | NLM W2 A N148vb no.183 2019 | DDC 617.6/010727--dc23
LC record available at https://lccn.loc.gov/2019019953

For sale by the U.S. Government Printing Office
Superintendent of Documents
Mail Stop: SSOP
Washington, DC 20402–9328
Printed on acid-free paper.

NATIONAL CENTER FOR HEALTH STATISTICS

Vital and Health Statistics

Series 2, Number 183 April 2019

Data Quality Evaluation of the Dental Fluorosis Clinical Assessment Data From the National Health and Nutrition Examination Survey, 1999–2004 and 2011–2016

Data Evaluation and Methods Research

U.S. DEPARTMENT OF HEALTH AND HUMAN SERVICES
Centers for Disease Control and Prevention
National Center for Health Statistics

Hyattsville, Maryland
April 2019
DHHS Publication No. 2019–1383

Contents

Text Figures

Contents—Con.

Detailed Tables

Data Quality Evaluation of the Dental Fluorosis Clinical Assessment Data From the National Health and Nutrition Examination Survey, 1999–2004 and 2011–2016

by Division of Health and Nutrition Examination Surveys, National Center for Health Statistics, Centers for Disease Control and Prevention; and Division of Oral Health, National Center for Chronic Disease Prevention and Health Promotion, Centers for Disease Control and Prevention

Abstract

Introduction

This report provides a technical review of the 1999–2004 and 2011–2016 National Health and Nutrition Examination Survey (NHANES) dental fluorosis clinical assessment data.

Methods

Dental examiners conducted fluorosis assessments on participants at the mobile examination center, scoring up to 28 teeth per participant according to Dean's Fluorosis Index (DFI). Person-level values were assigned based on the value for the lesser of the two most affected teeth to determine fluorosis prevalence, overall and by severity. Data quality was evaluated by comparing dental and reference examiners' repeat examinations to determine how consistently examiners rated the extent of fluorosis. Kappa statistics and percent agreement were calculated to assess this. Additionally, a synthetic cohort of youth aged 6–9 years in 2001–2004 and 16–19 years in 2011–2014 was analyzed to determine if the percentage with fluorosis was constant with age, as would be expected given that fluorosis develops before teeth erupt.

Results

Weighted kappa statistics comparing person-level scores between dental and reference examiners across different weighting schemes for five dental examiners in 1999–2004 ranged from 0.51 to 0.87 and for four examiners in 2011–2016 from 0.60 to 0.98. Weighted kappa statistics comparing person-level scores for the same participant by the same examiner in 1999–2001 ranged from 0.56 to 0.72. Percent agreement between dental and reference examiners in defining very mild or greater fluorosis was 88.8% in 1999–2004 and 89.4% in 2011–2016; for moderate or severe, agreement was 97.1% and 94.4%, respectively. Variability in fluorosis prevalence estimates by DFI category was seen across 2-year NHANES cycles. Adjusted prevalence of mild or greater fluorosis in first permanent molars for the synthetic cohort was 9.5% (standard error [SE] = 0.01) in youth aged 6–9 years in 2001–2004 but 46.9% (SE = 0.04) in youth aged 16–19 years in 2011–2014.

Conclusions

Two-year fluorosis prevalence estimates by DFI category demonstrated variability within and across the 6-year time periods. Inter-examiner reliability statistics found that agreement ranged from 0.51 to 0.98 for the nine primary dental examiners. These values indicate moderate (0.41–0.60) to almost-perfect (0.81–0.99) agreement based on Landis and Koch, and mostly adequate agreement (33 out of 36 kappas ≥ 0.6) based on McHugh. There was also high percent agreement in defining very mild or greater fluorosis. The proportion of the total number of examinations for which a gold standard examination was conducted, however, was low and the DFI scoring method has high examiner subjectivity. The observed increase in dental fluorosis prevalence with age between 2001–2004 and 2011–2014, based on the synthetic cohort analyses, is not biologically plausible. This suggests that there may have been some change in the way the examiners evaluated the level of fluorosis over time. The quality assessment findings in this report should be strongly considered when determining whether these data are appropriate for the user's analytic objectives, including studies of prevalence and trends.

Keywords: enamel • fluoride • NHANES

Introduction

Fluoride is a normal component of teeth and is incorporated into developing dental enamel before a tooth erupts into the mouth. When fluoride is present in saliva and dental plaque, it inhibits bacterial by-products (i.e., acids) that dissolve the hard surfaces of the teeth and enhances the remineralization of the tooth enamel (1–4).

In many U.S. communities, fluoride is increased in the water systems to reach recommended levels for dental caries prevention. Fluoride is also included in a number of commercial products, including toothpaste, mouth rinse, some bottled water, dietary supplements, and professionally applied or prescribed dental gel, foam, and varnish (2,5).

Fluoride exposure from any source during the period of tooth development can lead to dental (or enamel) fluorosis (4,5). Dental fluorosis is characterized by an increasing porosity or hypomineralization of the tooth enamel that leads to visual changes of the enamel that appear after a tooth erupts (6,7). The severity of dental fluorosis depends on the dose and duration of fluoride ingestion during tooth development (7,8). With very mild and mild fluorosis, there are small, opaque, paper white areas scattered across the tooth's enamel. With moderate fluorosis, more than one-half of the enamel surface is opaque white. In the severe form, the affected tooth's surface becomes pitted and brittle. In contrast, teeth unaffected by fluorosis have a smooth, glossy appearance and are usually a pale, creamy, white color (6–10).

Tooth enamel is no longer susceptible to the developmental effects of fluoride after the tooth's pre-eruptive development is complete (1). Therefore, concerns regarding dental fluorosis risk are limited to children from birth until about 8 years of age (1,2,9,10). By age 8, with the exception of permanent third molars, there is no further risk of fluorosis because the enamel of the permanent teeth is fully mineralized (11). Additionally, by age 6, the swallowing reflex has developed sufficiently for most children, so inadvertent swallowing of fluoride toothpaste or mouth rinse is less likely.

In the United States, dental fluorosis is generally considered a cosmetic effect with no negative functional effect (2,5,12). The severe form of dental fluorosis, however, may have adverse dental effects because the pitting can compromise the protective function of the enamel and the affected area can break away, resulting in excessive wearing of the teeth (2,7,11,13). However, severe fluorosis is rare among U.S. youth and requires fluoride exposure at high levels (2 mg/L or more) for possibly long durations (13).

The National Health and Nutrition Examination Survey (NHANES) collected dental fluorosis clinical assessment data from 1999 through 2004 and from 2011 through 2016 as part of the NHANES oral health component. Dental fluorosis clinical data were also collected during 2009–2010; however, there were some key differences with fluorosis assessment in 2009–2010 compared with the other years, including assessment by dental hygienists instead of dentists and assessment on only six maxillary anterior teeth compared with all permanent teeth. These data are not described further as part of this quality assessment but will be evaluated separately before data release. The objective of this report is to provide a technical review of the NHANES dental fluorosis clinical assessment data collected from 1999 through 2004 and from 2011 through 2016. First, the dental fluorosis data collection and quality assurance and control procedures are described. Second, an evaluation of the data quality by examining rater variability and reliability is conducted. Third, the biological plausibility of the prevalence estimates for youth is evaluated. The potential impact of the NHANES complex sample design on computed estimates is also discussed.

Methods

NHANES Oral Health Data

NHANES is a cross-sectional survey designed to monitor the health and nutritional status of the civilian, noninstitutionalized U.S. population. For more information on NHANES, please visit: https://www.cdc.gov/nchs/nhanes/index.htm. The survey consists of interviews conducted in participants' homes and standardized physical examinations in mobile examination centers (MEC), including oral health assessments. The NHANES protocol is approved by the National Center for Health Statistics Research Ethics Review Board. All adult participants provide written informed consent. A parent or guardian provides consent for participants under 18 years of age. Additionally, children aged 7–11 years provide assent and those aged 12–17 years also provide consent. Details of NHANES operations and sample designs are published elsewhere (14–16).

The NHANES oral health component is currently the only source of nationally representative, clinical oral health data. With the exception of NHANES II (1976–1980), oral health data have been collected since the first National Health Examination Survey in 1959. Over the years, NHANES has collected data on oral health topics such as dental caries, periodontal disease, sealants, tooth loss, gingivitis, soft tissue lesions, tooth erosion, occlusion, traumatic injuries, and dentures. From 1999 through 2004 and 2011 through 2016, dental fluorosis clinical examination data were also collected at the MEC by certified dentists on all permanent teeth excluding third molars. During 2009–2010, dental fluorosis clinical examination data were also collected by dental hygienists on six maxillary anterior teeth. Reports providing an overview of the NHANES oral health component and data quality control procedures have been previously published (17–19).

Sample surveys such as NHANES, use a systematic sampling of people to estimate the prevalence of a condition or characteristic of the population instead of studying the entire population. NHANES sample persons are selected through a four-stage, probability cluster sampling process, which includes primary sampling units composed of counties in the first stage, clusters of housing units within the counties in the second stage, households in the third stage, and persons within households in the fourth stage. During the first stage of sampling for NHANES, about 15 primary sampling units (i.e., counties) are selected annually from approximately 3,000 U.S. counties. Ultimately, about 10,000 persons are interviewed and examined from about 30 locations across the country during a 2-year survey cycle.

Certain subpopulations are oversampled to increase the precision of estimates computed on them. During NHANES 1999–2004, Mexican-American and black persons, low-income white and other racial and ethnic persons, adolescents aged 12–19 years, and adults aged 70 and over were oversampled. During NHANES 2011–2016, non-Hispanic black, non-Hispanic Asian, and Hispanic persons were oversampled, in addition to low-income white and other racial and ethnic persons, and adults aged 80 and over. Adolescents were no longer oversampled.

Use of the survey sample weights in data analysis will account for the specific sample designs and the differential probabilities of selection (i.e., the oversampling). The sample weights can be considered a measure of the number of persons in the target population represented by the sampled person. Therefore, weighting the data using the appropriate sample weights for the specific survey years will produce estimates representative of the U.S. population.

It is important to note that NHANES is not designed to produce regional or subregional estimates, and data are released in 2-year survey cycles with the primary goal of providing national estimates with an adequate degree of precision (20). Due to the limited number of locations NHANES visits during a 2-year survey cycle, for measures that may have considerable geographic variation, it is recommended that analysts use combinations of 2-year cycles to improve the statistical reliability and stability of these estimates with larger variances (21).

NHANES Dental Fluorosis Clinical Assessment

During NHANES 1999–2004 and 2011–2016, the dental fluorosis clinical assessment was part of a larger oral health exam, which also included assessment of dentition (tooth count, dental caries, and dental sealants) and periodontal disease. From 1999 through 2004, all participants aged 2 years and over were eligible for the oral health examination. From 2011 through 2016, all participants aged 1 year and over were eligible. Assessment for a specific subcomponent of the examination was dependent on the participant's age. The age group eligible for the dental fluorosis clinical assessment ranged from 6–49 years in 1999–2004, 6–19 years in 2011–2013, and 6–29 years in 2014–2016.

The dental fluorosis assessment occurred after the tooth count and dental caries and sealant assessments, with the participant seated in a dental chair in a recumbent position and the dental examiner seated behind the participant.

The details of the NHANES dental fluorosis clinical assessment protocol for 1999–2004 and 2011–2016 are documented in the NHANES Dental Examiners and Oral Health Examiners Procedures Manuals, available from: https://wwwn.cdc.gov/nchs/data/nhanes/2003-2004/manuals/DentalExaminers-2004.pdf and https://wwwn.cdc.gov/nchs/data/nhanes/2015-2016/manuals/2016_Oral_Health_Examiners_Procedures_Manual.pdf.

Dental Examiners

During 1999–2004 and 2011–2016, the oral health examinations were conducted by dental examiners, who were dentists (D.D.S. or D.M.D.) licensed in at least one U.S. state. A health technician assisted in entering all examiner observations directly into a computerized data collection system at the MEC. During each survey cycle, at least two primary dental examiners performed the majority of dental fluorosis assessments, along with one backup dental examiner.

Examiners received an initial 5-day training, which consisted of lecture, model review, practice simulations, and standardization sessions. There was also a session where the examiner was "calibrated" to the reference examiner before the beginning of the survey. Following successful initial training, an examiner received 3-day field training at the NHANES MEC. This training consisted of more practice simulations and standardization and calibration sessions. Examiners also received periodic field visits for monitoring and recalibration. During 1999–2004 and 2011–2016, a single reference examiner provided training to all dental examiners and conducted periodic site visits to observe staff members and operational procedures.

Fluorosis Clinical Assessment

During the 1999–2004 and 2011–2016 dental fluorosis clinical assessment, the examiners evaluated all fully erupted permanent teeth, excluding third molars, using a dental mirror. In 1999–2004, examiners used a surface reflecting mirror and a No. 23 explorer for the assessment. In 2011–2016, only the mirror was used. Teeth were not dried with air before assessment.

The assessment proceeded tooth by tooth in a similar manner as the dental caries assessment, beginning with the maxillary right central incisor and proceeding posteriorly to the upper second molar. The same sequence was then repeated for the upper left, lower left, and lower right quadrants of the mouth.

Each tooth was scored according to Dean's Fluorosis Index (DFI) and assigned one of the DFI disease severity categories (Table 1) based on the area of the tooth surface with visible fluorosis and presence of pitting: normal (code 0), questionable (code 5), very mild (code 1), mild (code 2), moderate (code 3), or severe (code 4) (6,8). Missing teeth, deciduous (primary) teeth, permanent teeth not fully erupted, and teeth in which more than one-half of the visible surface area was obscured by a restoration, caries, or an orthodontic appliance were not assessed. These teeth were coded as cannot be assessed (code 9). Teeth may present with nonfluoride opacities, which resemble fluorosis but are due to other causes (22). A tooth having nonfluoride opacity was coded 8. The *NHANES Oral Health Examiners Manual* provides details on assessing a tooth for fluorosis and includes a table to help examiners differentiate between fluorosis and nonfluoride opacity. For example, because dental fluorosis is bilaterally distributed, examiners were told to observe the enamel condition of the tooth being assessed as well as the corresponding bilateral tooth. If the bilateral tooth relatively exhibited comparable enamel opacities or anomalies, then a fluorosis value was called to the recorder for the initially examined tooth.

DFI is the most widely used classification system for dental fluorosis in the United States and has been in use for more than 50 years. Differentiating between contiguous scoring categories in DFI demands that, based on visual examination only, the examiner assess the whole tooth, add all affected areas to determine the total tooth surface area affected, and decide whether that area is less than 25%, 25% to less than 50%, or 50% or greater of the total tooth surface (1,7). Assessing the area affected can be particularly challenging in posterior teeth due to the shorter cervico-cuspal height and the larger mesiodistal width dimensions compared with anterior teeth. Distinguishing fluorosis from other enamel defects can also be challenging for an examiner. Additionally, the questionable category was intended to be used when the examiner was unsure if the enamel opacity was normal or very mild. This may also create some confusion in assigning DFI values.

Quality Assurance Procedures

The specific quality assurance (QA) procedures for the oral health component are documented in detail and publicly available from: https://wwwn.cdc.gov/nchs/data/nhanes/2015-2016/manuals/2016_Oral_Health_Examiners_Procedures_Manual.pdf.

During 1999–2004 and 2011–2016, the reference examiner visited the MEC approximately twice a year to observe field operations and to replicate a certain number of random examinations by each examiner. The purpose of these gold standard examinations was to determine if the dental examiners were maintaining the examination standards achieved during training and were still calibrated to the reference examiner. The reference examiner determined if retraining and future monitoring of the dental examiner were needed. The reference examiner also conducted an annual retraining session for all dental examiners to reinforce existing protocols and to introduce protocol updates as needed. In addition, during 1999–2001, approximately 10% of examined participants were asked to return for a repeat examination. The purpose of these repeat examinations was to monitor internal consistency within examiners regarding the data collection process. No repeat examinations were conducted from 2011 through 2016.

Since 1999, data for the oral health examination have been recorded directly into a computerized data collection system at the NHANES MEC. The system is integrated centrally and allows for ongoing monitoring of much of the data. As part of the quality control procedures, all data are reviewed systematically for logical inconsistencies. Before data release, the collected data are further reviewed.

Response Rates and Completion Rates

The NHANES examination response rate for the 2-year cycles for youth aged 6–19 years (the common age range across all survey years) ranged from 83.6% to 86.1% during 1999–2004 and 64.7% to 76.8% during 2011–2016 (https://wwwn.cdc.gov/nchs/nhanes/ResponseRates.aspx#response-rates). Table 4 provides the number and percentage of participants aged 6–19 years who were MEC examined and who were eligible for the dental fluorosis clinical assessment by age and race and Hispanic-origin group. Among participants aged 6–19 years who were MEC examined, 96.0% (n = 9,826) and 98.1% (n = 7,601) were eligible for the dental fluorosis assessment in 1999–2004 and 2011–2016, respectively.

Data Evaluation

All statistical analyses presented in this report were conducted separately by two data analysts using the SAS System for Windows, version 9.4 (SAS Institute, Inc., Cary, N.C.) and SUDAAN, version 10.0 (RTI International, Research Triangle Park, N.C.). Some analysis was also conducted using Stata, version 13.1 (StataCorp, College Station, T.X.).

All analyses were conducted at the person level. A person was assigned a dental fluorosis severity value based on the lesser of the two most affected tooth-level DFI values based on up to 28 assessed teeth. Tooth-level DFI values of cannot be assessed (code 9) and nonfluoride opacity (code 8) were recoded to missing. The questionable category, which was recorded during data collection as 5, was recoded to 0.5. This is consistent with Dean's assignment of values, and reflects a questionable value falling between normal and very mild, in terms of severity. Only youth with at least two teeth with a non-missing DFI value were assigned a person-level DFI and included in further analyses. Table 2 shows how person-level DFI values were assigned based on the tooth-level DFI values.

Results

Rater Variability

As described previously, several QA processes were implemented during the dental fluorosis clinical assessment. In addition, evaluation of rater variability and reliability was assessed on the final data. Specifically, percent agreement and kappa statistics were calculated to assess intra- and inter-examiner reliability. The kappa statistic measures agreement between two examiners or repeat examinations by the same examiner, taking into account agreement that would be expected based on chance and the true percentage of the population in each category. To maintain consistency with previous NHANES oral health data quality reports, the guidelines suggested by Landis and Koch (23) were used for the qualitative interpretation of numeric kappa statistic value ranges as less than chance agreement (< 0), slight agreement (0.00–0.20), fair agreement (0.21–0.40), moderate agreement (0.41–0.60), substantial agreement (0.61–0.80), and almost-perfect agreement (0.81–0.99) (Table 3). Proposed by McHugh (24), any kappa below 0.60 may also be used as an indicator of inadequate agreement among raters.

Intra-examiner reliability: Evaluation of repeat examinations, 1999–2001

During 1999–2001, approximately 10% of examined participants aged 6–49 years had a repeat fluorosis examination that occurred days later. Intra-examiner reliability was assessed for each dental examiner by comparing results of the examiner's initial and repeat examinations for the same participant using the kappa statistic. Details on these repeat examinations were previously published (17). To summarize, the weighted kappa statistics comparing the DFI values for the same participant by the same examiner ranged from 0.56 to 0.72 across the three main dental examiners who conducted exams from 1999 through 2001. Based on the guidelines by Landis and Koch (23), this range of kappa values is considered moderate to substantial agreement. The difference in DFI values for the same teeth assessed in the same way days apart, however, indicates the subjective nature of the DFI scoring method.

Inter-examiner reliability: Evaluation of gold standard examinations, 1999–2004 and 2011–2016

Gold standard examinations conducted during 1999–2004 and 2011–2016 allowed assessment of inter-examiner reliability. This included computing percent agreement and kappa statistics comparing person-level DFI values computed by the reference examiner and the dental examiner for the same participant. Gold standard examinations were conducted on 3.6% (*n* = 356) and 2.8% (*n* = 210) of participants aged 6–19 years who received the dental fluorosis examination in 1999–2004 and 2011–2016, respectively. Complete data (i.e., at least two teeth with valid DFI values for both examiner and reference) for this analysis were available for 339 participants aged 6–19 years in 1999–2004 and 198 in 2011–2016. These analyses did not use the survey examination weights.

The simple unweighted kappa assigns a weight of 1 for observations in perfect agreement between the two examiners and 0 for any disagreement (Table 5). However, disagreement between each category may not be of equal importance, such as a difference between very mild and mild compared with a difference between very mild and severe. Therefore, four different weighting schemes, specified in Table 5, were used to assign various weights when the two examiners disagreed by one or more DFI levels. All kappa analyses were computed on person-level DFI values using the values of 0 (normal), 0.5 (questionable), 1 (very mild), 2 (mild), 3 (moderate), and 4 (severe).

The first weighting scheme used the Cicchetti-Allison method, which is the default method for computing weighted kappa statistics in SAS (25). Given ordered scores $S_1, S_2 \ldots S_r$, the Cicchetti-Allison weights are defined as:

$$1 - \frac{|s_j - s_i|}{|s_r - s_1|}.$$

For example, given the ordered DFI values of 0, 0.5, 1, 2, 3, and 4, the Cicchetti-Allison weight applied when the dental examiner's DFI value for a participant was 0.5 (questionable) and the reference examiner's DFI value for the same participant was 1 (very mild) would be

$$1 - \frac{(1 - 0.5)}{(4 - 0)}$$

or 0.875 (see Table 5 for the full Cicchetti-Allison weighting matrix).

The second weighting scheme used the Fleiss-Cohen method (26). Given the same ordered scores $S_1, S_2 \ldots S_r$, the Fleiss-Cohen weights are defined as:

$$1 - \frac{(s_j - s_i)^2}{(s_r - s_1)^2}.$$

For the same scenario when the dental examiner's DFI value for a participant was 0.5 (questionable) and the reference examiner's DFI value for the same participant was 1 (very mild), the Fleiss-Cohen weight would be

$$1 - \frac{(1 - 0.5)^2}{(4 - 0)^2}$$

or 0.984375 (see Table 5 for the full Fleiss-Cohen weighting matrix).

The third and fourth weighting schemes applied custom weights based on previously published literature and were implemented using Stata, as user-defined weights cannot be specified directly in SAS. The first custom scheme assigned a weight of 1 for perfect agreement, 0.667 for one category difference, 0.33 for two category differences, and 0 for all other differences (Table 5) regardless of the actual category values. The second custom scheme (27) assigned a weight of 1 for perfect agreement, 0.5 for one category difference, and 0 for all other differences (Table 5) regardless of the actual category values.

Table 6 provides the percent agreement and the kappa statistics between the dental examiners and the reference examiner on the calculated person-level DFI for the available gold standard observations by 2-year survey cycle. Table 7 summarizes these data by aggregated 6-year survey period.

The different weighting schemes show how the assignment of weights changed the overall kappa statistics. The simple unweighted kappa was the most stringent, with a weight of 1 for perfect agreement and no weight for any disagreement. This led to examiner kappa values ranging from 0.35 to 0.78 during the two 6-year time periods (Table 7). When a weight of 1 was applied for perfect agreement, but 0.5 was allowed for one category disagreement (Custom #2), kappa values ranged from 0.51 to 0.84. The Custom #1 scheme with weights of 1, 0.66, and 0.33, which had kappa values ranging from 0.57 to 0.87, was closest to the SAS default Cicchetti-Allison weights, and resulted in kappa values ranging from 0.62 to 0.93. The Fleiss-Cohen weights led to the highest kappa values, which ranged from 0.77 to 0.98. As expected, all weighted kappas resulted in improved agreement between dental examiner and reference examiner when compared with the unweighted kappa results.

When considering the results across all four weighting schemes presented in Table 7, the agreement between the dental examiners and reference examiner ranged from 0.51 to 0.87 for the five primary dental examiners who conducted examinations in 1999–2004 and from 0.60 to 0.98 for the four primary examiners during 2011–2016. These reliability statistics would be considered moderate to almost-perfect agreement based on the Landis and Koch interpretation of kappa values (Table 3). When McHugh's (24) requirement of a kappa value of at least 0.60 for adequate agreement is applied, the majority of the kappa values indicate adequate agreement; that is, 33 out of the 36 kappa values (computed for nine examiners for four different weighting schemes) were at or above 0.60 (Table 7). Based on the default SAS Cicchetti-Allison weighting scheme, for the 2-year cycles, agreement was greater than 0.60 for all but three examiners who had kappa values ranging from 0.50 to 0.57 (Table 6); for the two 6-year periods, values ranged from 0.62 to 0.93 (Table 7). Based on the Fleiss-Cohen weighting scheme, all 2- and 6-year kappa values for all nine examiners were above 0.60 (0.65–0.98 and 0.77–0.98, respectively).

If prevalence within each fluorosis severity category is important, then the unweighted rather than weighted kappas should be considered. In this case, unweighted kappa values for examiners ranged from 0.35 to 0.78, with five of nine examiners having values below 0.60 when all of an examiner's gold standard data across a 6-year period were analyzed (Table 7).

Percent Agreement Between Dental and Reference Examiners in Categorizing Fluorosis

For prevalence studies, the main outcome is often the proportion of persons with any fluorosis, regardless of level of severity. In this case, reliably distinguishing between some and no fluorosis becomes important. Also, determining the presence of severe fluorosis has the greatest public health implications, but the prevalence of severe fluorosis is very rare in the United States (11,28). When DFI categories are collapsed into broader categories and fluorosis is defined as very mild or greater severity, the examiner and reference agreed on 88.8% of cases in 1999–2004; agreement was 89.4% in 2011–2016. When fluorosis is defined as moderate or severe, agreement was 97.1% in 1999–2004 and 94.4% in 2011–2016. It is expected that agreement would improve when broader categories are assessed since agreement is achieved as long as both examiners rate the same DFI category or higher.

As a means of detecting systematic scoring biases, differences between the dental examiner and reference examiner were further explored by examining the direction of the disagreement between the dental examiner's and the reference examiner's person-level DFI values (Table 8). Among the cases where there was disagreement between the examiner and reference, in 1999–2004, the examiner scored higher than the reference examiner on 48.8% of these cases, while the reference examiner scored higher than the dental examiner on 51.2% of cases. In 2011–2016, the examiner scored higher than the reference examiner on 54.8% of cases, whereas the reference examiner scored higher than the dental examiner on 45.2%. The majority of the disagreement was by one DFI level. Specifically, in 1999–2004, among the cases where there was disagreement, the dental examiner and reference examiner differed by one DFI level for 82.7% of these cases, by two levels for 15.0%, and by three levels for 2.4%. In 2011–2016, 88.7% of the cases with disagreement differed by one DFI level and 11.3% by two DFI levels.

Table 9 presents the percentage of participants aged 6–19 years classified at each DFI level among those with gold standard readings. Overall, for the two 6-year survey periods, there were similarities between the dental examiners and reference examiner in the person-level DFI values. In 1999–2004, 48.7% were classified by the dental examiner and 49.8% by the reference examiner with DFI values of very mild or greater. In 2011–2016, the percentages were 74.2% and 73.7%, respectively. It is important to note that the number of gold standard examinations and the percentage of gold standard examinations among all examinations were relatively low. However, it is also important to note that the dental fluorosis assessment is a subcomponent of a larger oral health examination, which is only one of a number of NHANES examinations. The need for additional examinations, including gold standard and repeat examinations, must be considered along with increased respondent burden and costs.

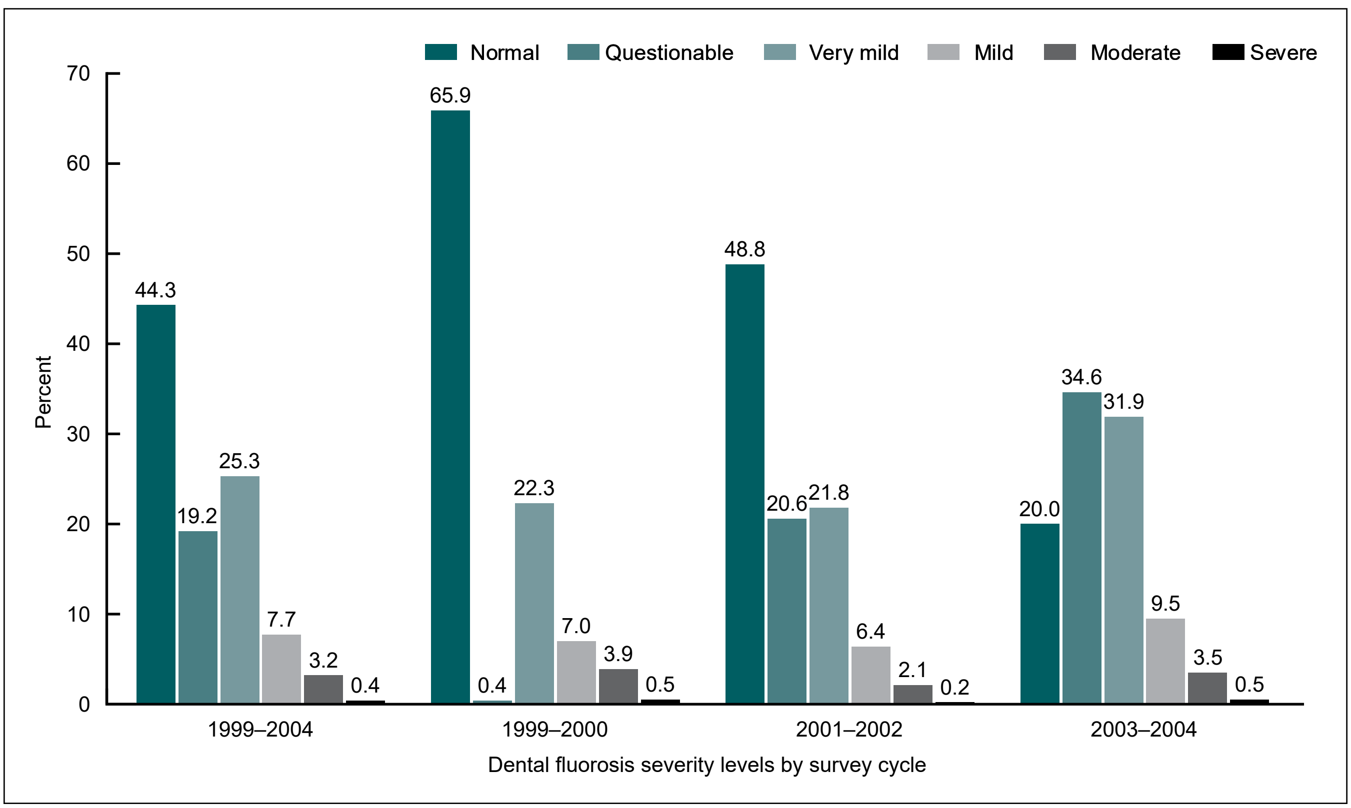
Prevalence of Dental Fluorosis Severity Among Youth

As stated previously, the goal of NHANES is to produce estimates representative of the civilian noninstitutionalized U.S. population. Prevalence estimates were computed from person-level DFI values for youth aged 6–19 years for any fluorosis, which included youth with very mild, mild, moderate, or severe fluorosis. Additional analysis is also presented on youth with only moderate or severe fluorosis. Since person-level DFI is based on the lesser of the two most affected teeth, only youth who had at least two teeth with a non-missing DFI value were further available for analysis (n = 9,395 for 1999–2004 and n = 7,158 for 2011–2016) (Table 4). Examination sample weights were used to account for differential probabilities of selection, nonresponse, and noncoverage. Taylor series linearization was used to calculate standard errors to account for the complex sampling design.

Table 10 provides weighted 2- and 6-year estimates of dental fluorosis severity among youth aged 6–19 years (the common age range across all years) for 1999–2004 and 2011–2016. During 1999–2004, 25.3% of youth aged 6–19 years were estimated to have very mild fluorosis, 7.7% mild, 3.2% moderate, and 0.4% severe (Table 10, Figure 1). In 2011–2016, 35.6% of youth aged 6–19 years were estimated to have very mild fluorosis, 21.5% mild, 13.4% moderate, and 1.0% severe (Table 10, Figure 2). There were differences in the overall prevalence (i.e., very mild or greater) and in the specific severity categories between 1999–2004 and 2011–2016. Figure 3 shows how the cumulative distribution of DFI values differs between 1999–2004 and 2011–2016. In 1999–2004, 3.6% of youth aged 6–19 years had moderate or severe fluorosis, whereas in 2011–2016, the percentage increased to 14.4%.

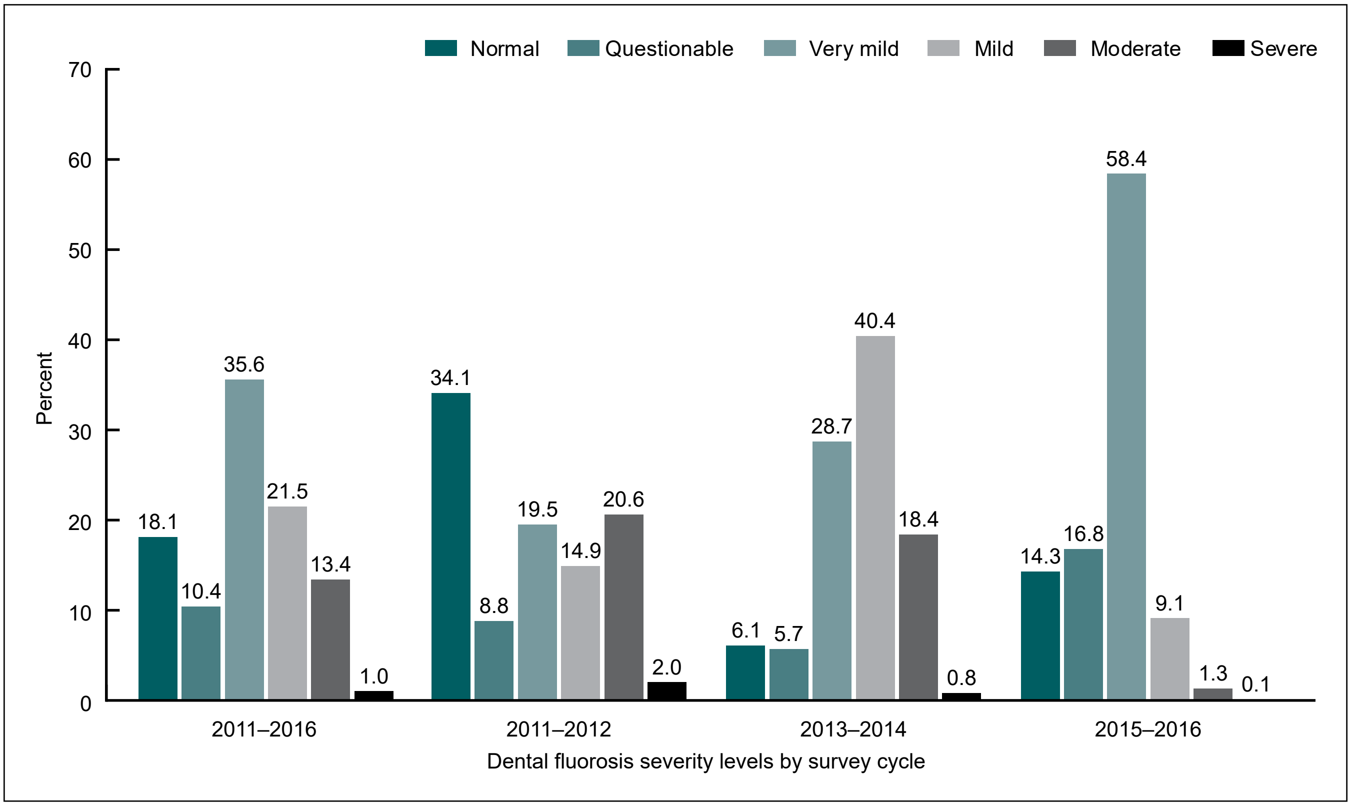
Variability in the prevalence of different dental fluorosis severity levels was seen within each 6-year time period, as well as across the two time periods (Figures 1, 2, and 4). For example, during 1999–2004, questionable fluorosis ranged from 0.4% to 34.6%. During 2011–2016, mild fluorosis prevalence ranged from 9.1% to 40.4%, and moderate fluorosis prevalence ranged from 1.3% to 20.6% (Table 10).

Figure 1. Percentage (weighted) dental fluorosis severity levels based on person-level Dean's Fluorosis Index among youth aged 6–19 years: National Health and Nutrition Examination Survey, 1999–2004



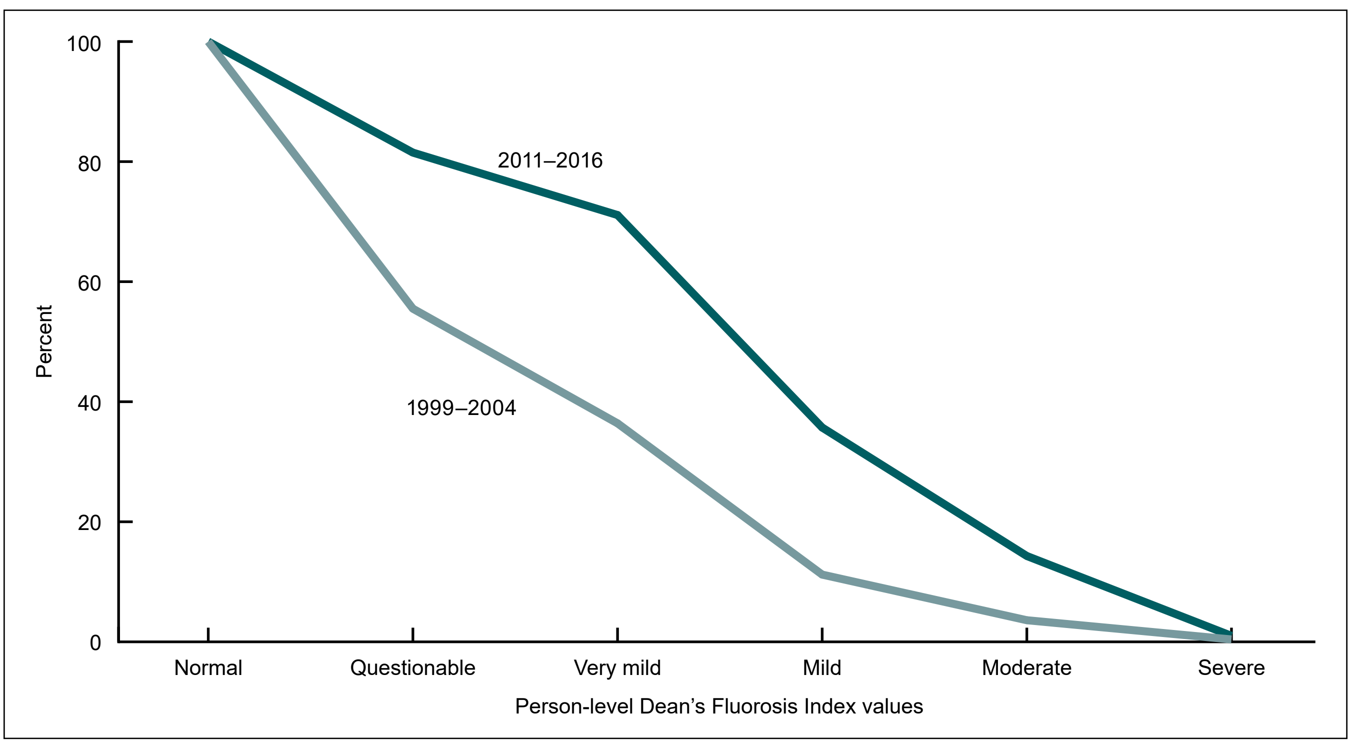
SOURCE: NCHS, National Health and Nutrition Examination Survey, 1999–2004.

Figure 2. Percentage (weighted) dental fluorosis severity levels based on person-level Dean's Fluorosis Index among youth aged 6–19 years: National Health and Nutrition Examination Survey, 2011–2016



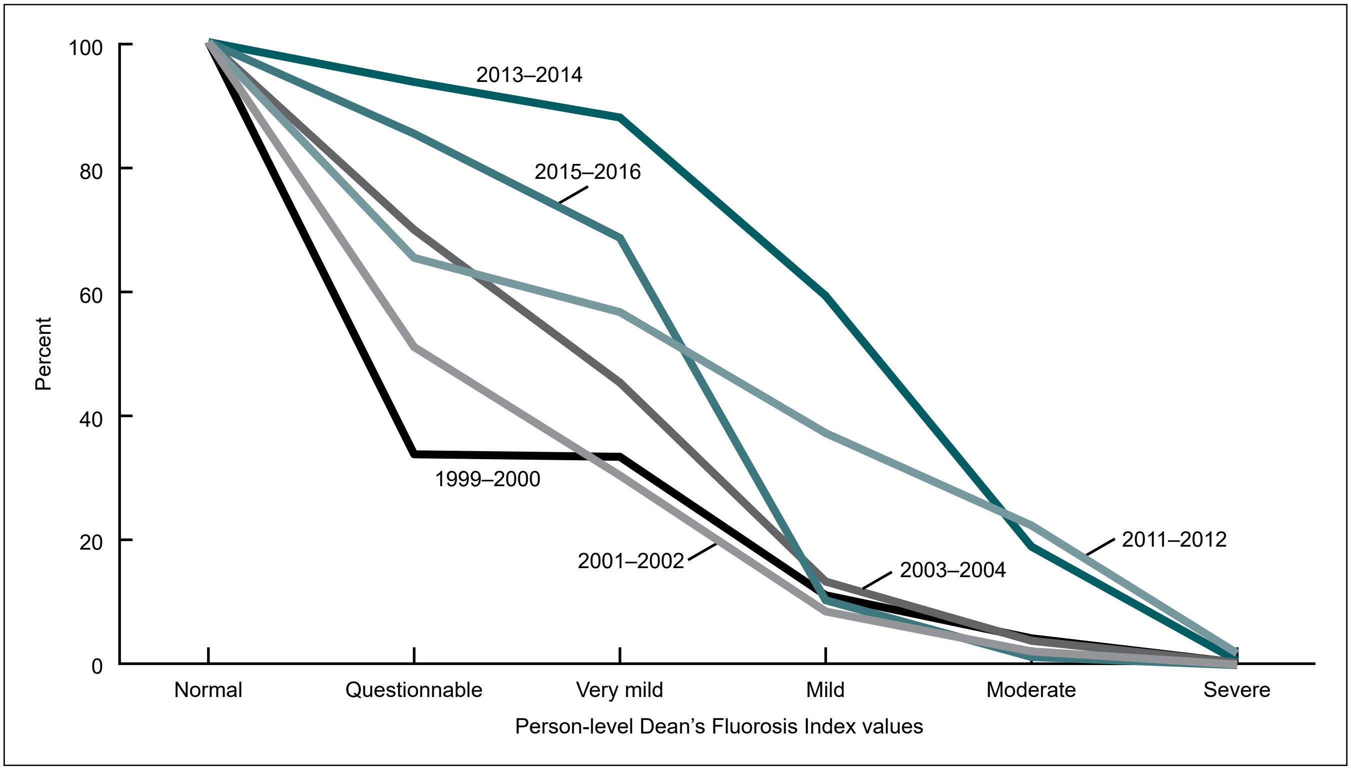
SOURCE: NCHS, National Health and Nutrition Examination Survey, 2011–2016.

Figure 3. Cumulative distribution of dental fluorosis severity levels, based on person-level Dean's Fluorosis Index values, for youth aged 6–19 years, by 6-year National Health and Nutrition Examination Survey cycles: 1999–2004 and 2011–2016



SOURCE: NCHS, National Health and Nutrition Examination Survey, 1999–2004 and 2011–2016.

Figure 4. Cumulative distribution of dental fluorosis severity levels, based on person-level Dean's Fluorosis Index values, for youth aged 6–19 years, by 2-year National Health and Nutrition Examination Survey cycles: 1999–2004 and 2011–2016



SOURCE: NCHS, National Health and Nutrition Examination Survey, 1999–2004 and 2011–2016.

Assessing Biological Plausibility of Prevalence Estimates

Further evaluation of the NHANES 1999–2004 and 2011–2016 dental fluorosis clinical assessment data was conducted to determine whether the prevalence estimates presented above could be consistent with the known etiology of fluorosis. Fluoride ingestion before tooth eruption is the only known cause of fluorosis (7); therefore, there should be no change in fluorosis prevalence among erupted permanent teeth of the same tooth type in the same birth cohort over time. For this analysis, a synthetic birth cohort that included youth aged 6–9 years in 2001–2004 (born 1992–1998; *n* = 1,097) and youth aged 16–19 years in 2011–2014 (born 1992–1998; *n* = 1,193) was constructed. Youth aged 6–9 years captured a broader age range for complete tooth eruption. It was hypothesized that the prevalence of dental fluorosis in the first permanent molars (which usually appear between 6 and 7 years) (29) in youth aged 6–9 years in 2001–2004 would not be different from the prevalence in youth aged 16–19 years in 2011–2014, given that fluorosis develops before teeth erupt. Only youth who had at least two teeth with a non-missing DFI value were included in the analysis. Logistic regression and computed predictive marginals were used to estimate the prevalence of mild or greater and moderate or severe dental fluorosis in the first permanent molars (identified in the data file as tooth numbers 3, 14, 19, and 30) among youth aged 6–9 years in 2001–2004 and 16–19 years in 2011–2014. Adjusted models, which included sex and race and Hispanic origin, were run to control for possible population changes over the time period.

Prevalence (adjusted for age and race and Hispanic origin) of mild or greater fluorosis in the first permanent molars of youth aged 6–9 years in 2001–2004 was 9.5% (standard error [SE] = 0.01) and among youth aged 16–19 years in 2011–2014 was 46.9% (SE = 0.04, $p < 0.001$). This increase between surveys was unexpected, because the prevalence of fluorosis in the first permanent molar should not change over time among a similar birth cohort given that tooth eruption has already occurred by 6–9 years of age. Similarly, adjusted prevalence of moderate and severe fluorosis in the first permanent molars of the birth cohort increased from 2.8% (SE = 0.01) in 2001–2004 to 17.7% (SE = 0.03) in 2011–2014 ($p < 0.001$). Estimates were similar in unadjusted models.

Impact of Oversampling in NHANES

During 1999–2016, there were changes in the racial and ethnic composition of the U.S. population. In 2011–2016, NHANES started oversampling non-Hispanic Asian American persons and continued the oversampling of Hispanic persons, which began in 2007. The oversampling of adolescents was also discontinued in 2007. It may be possible that underlying changes in the population composition of the United States or NHANES sample design changes affected the overall prevalence estimates of dental fluorosis and contributed to some of the observed differences in estimates between 1999–2004 and 2011–2016.

Closer examination of the race and Hispanic-origin estimates from 2011–2016, however, showed no differences in the prevalence of dental fluorosis for the specific severity categories across the different race and Hispanic-origin groups (Figure 5). In 1999–2004, the race and Hispanic-origin groups (specifically, non-Hispanic white, non-Hispanic black, and Mexican American) generally followed a similar pattern to 2011–2016 (Figure 6).

When comparing 1999–2004 with 2011–2016, there is a very similar pattern to that shown in Figure 3, where for all categories of dental fluorosis, estimates are higher in 2011–2016 compared with 1999–2004 for non-Hispanic white, non-Hispanic black, and Mexican-American groups. This was also true when comparing estimates for youth aged 6–11 years and 12–19 years (Figure 7).

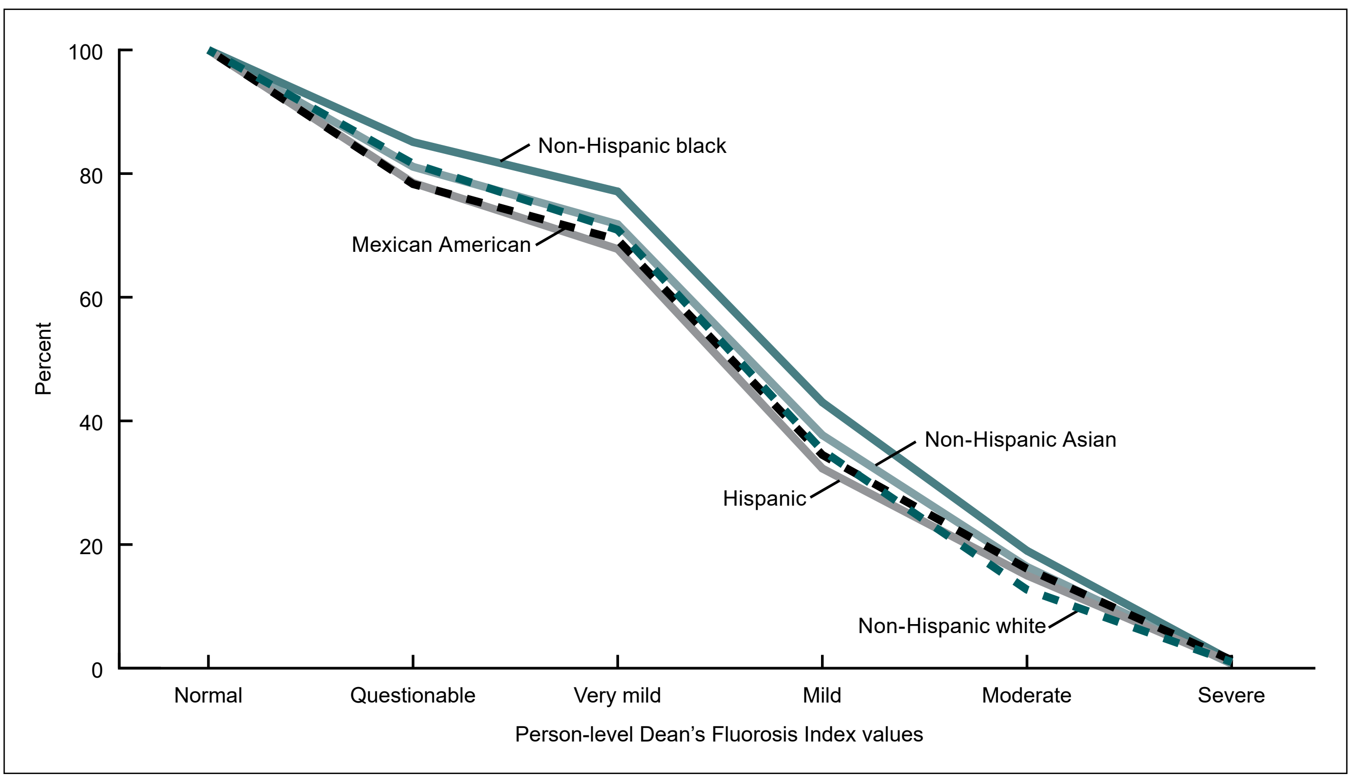
Impact of Geographic Variation in Water Fluoridation Levels

As mentioned earlier, in a given survey year, about 5,000 persons are selected from about 15 counties out of approximately 3,000 U.S. counties. Fluoride levels in drinking water are known to vary in communities across the country. Given this, the impact of geographic variation in water fluoridation levels was further examined.

The percentage of the U.S. population receiving fluoridated drinking water has increased over time (https://www.cdc.gov/fluoridation/statistics/fsgrowth.htm). Since 1962, the optimal fluoride concentration range in drinking water for community water systems is 0.7–1.2 mg/L, depending on the average maximum daily air temperature in the area (30). In 2015, the recommended maximum changed to 0.7 mg/L for all areas (2). Not all counties have fluoridated water systems and those that do have known variation in fluoride concentrations (https://nccd.cdc.gov/DOH_MWF/Default/WaterSystemDetails.aspx).

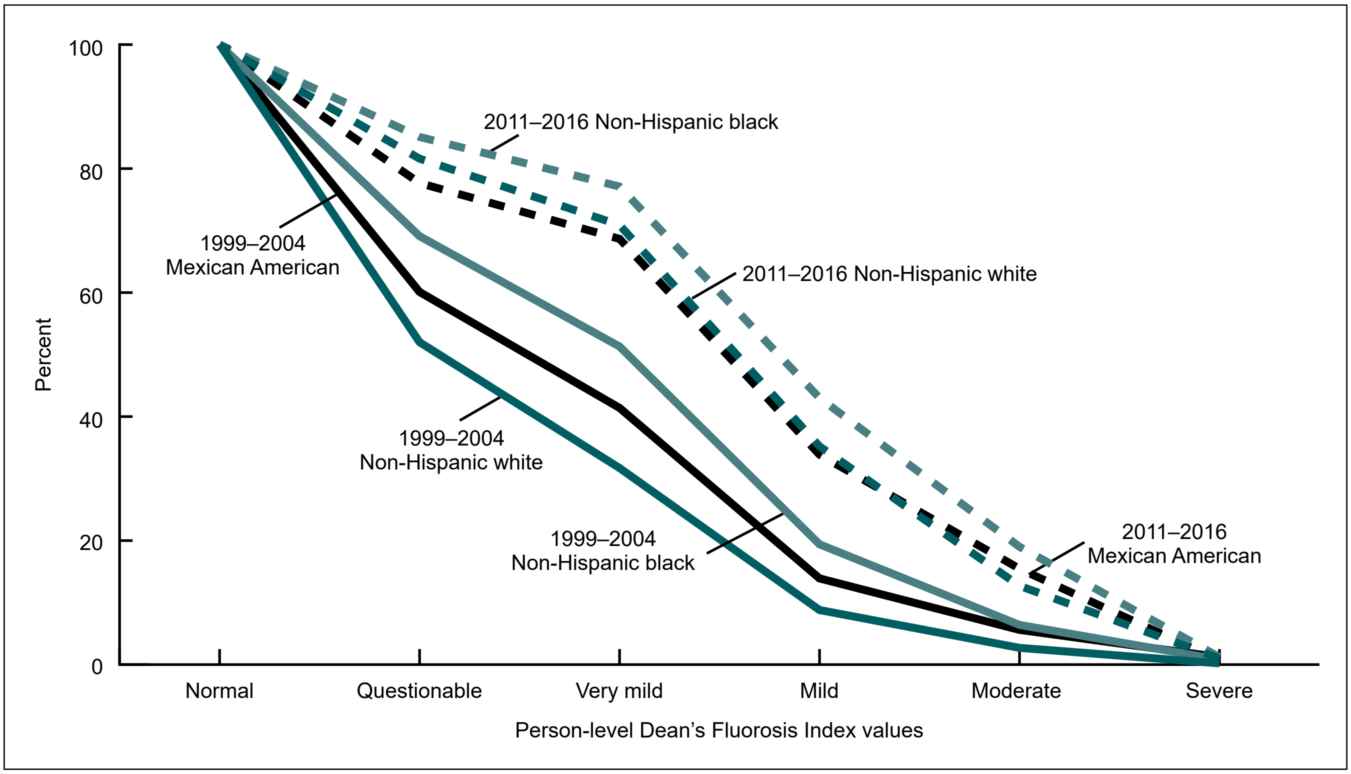
The Water Fluoridation Reporting System (WFRS) is a tool developed by the Centers for Disease Control and Prevention and the Association of State and Territorial Dental Directors that helps states manage the quality of their water fluoridation programs. WFRS collects community water fluoridation information for each Public Water System

Figure 5. Cumulative distribution of dental fluorosis severity levels, based on person-level Dean's Fluorosis Index values, for youth aged 6–19 years, by race and Hispanic origin: National Health and Nutrition Examination Survey, 2011–2016



SOURCE: NCHS, National Health and Nutrition Examination Survey, 2011–2016.

Figure 6. Cumulative distribution of dental fluorosis severity levels, based on person-level Dean's Fluorosis Index values, for youth aged 6–19 years, by race and Hispanic origin: National Health and Nutrition Examination Survey, 1999–2004 and 2011–2016



SOURCE: NCHS, National Health and Nutrition Examination Survey, 1999–2004 and 2011–2016.

(PWS), including information on whether a PWS has a naturally occurring fluoride concentration of 0.6 mg/L or greater, or a PWS has increased fluoride to adjust to the recommended level.

Determining the water fluoridation level in a given NHANES location is not straightforward, given that a county may include multiple water systems. For example, in 2013, there were 345 total PWSs for seven of the NHANES locations visited that year. Nine of the 345 PWSs had levels higher than 1.2 mg/L. These nine were all from a location with a substantial number of PWSs.

It may be possible that some geographic variation in local water fluoride levels across survey locations contributed to the differences observed in the percentage of dental fluorosis by examiner, as different examiners were assigned to different survey locations. However, each examiner's set of participants is not a nationally representative sample, so there would be expected variation by examiner.

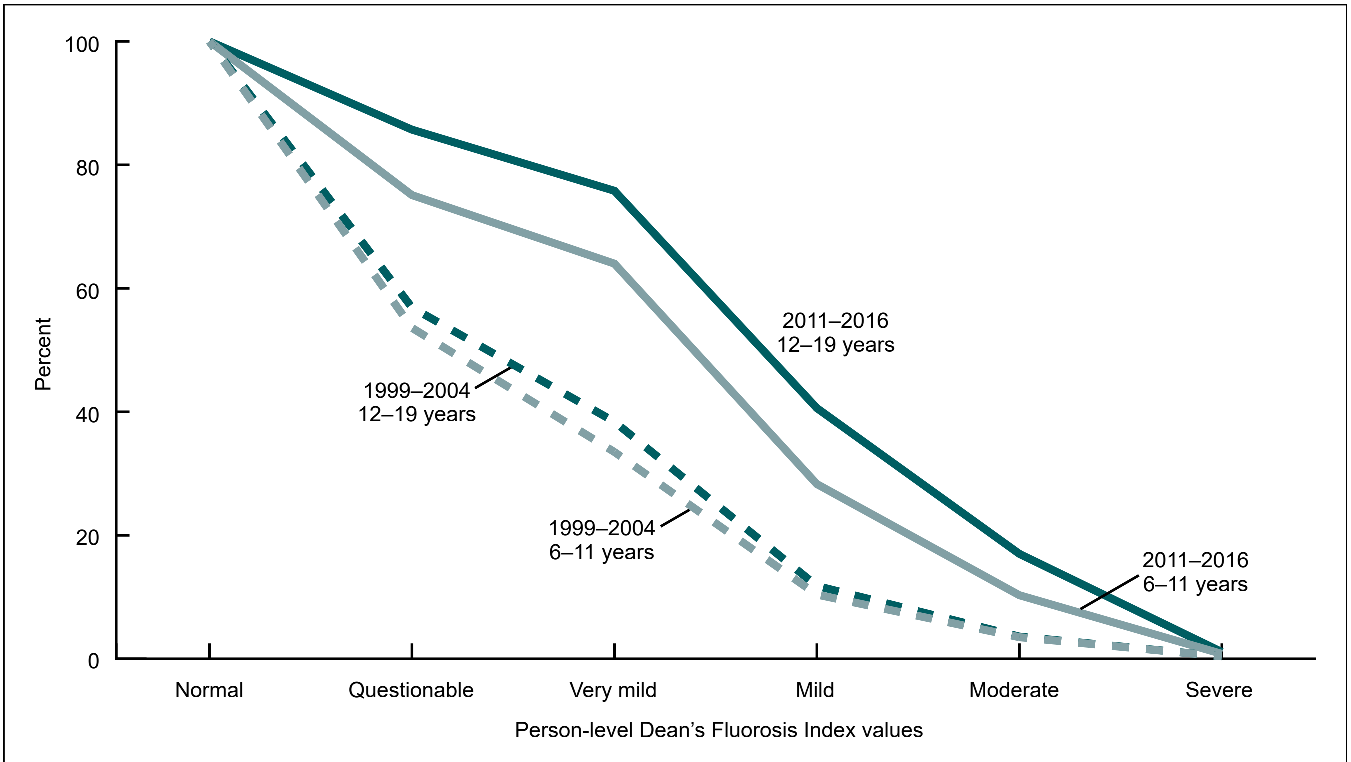
When examining dental fluorosis, it is important to note that the water fluoride level at the time of examination is not the water fluoride level during tooth formation. As stated earlier, dental fluorosis typically occurs from fluoride exposure during early tooth development before eruption. Water fluoride levels many years before the NHANES oral health examination would need to be assessed to determine exposure levels rather than levels for participants at the time of examination. For example, fluoride exposure for a 19-year-old participant examined in 1999 would have occurred during 1980–1986. Determining actual fluoride exposure during childhood, however, is further complicated by the uncertainty of migration patterns among families (i.e., if they ever moved, taking them from one PWS to another).

Mascarenhas (31) published a comprehensive literature review on risk factors for dental fluorosis and provided detailed information on studies discussing potential changes in fluoride ingestion based on fluoridated drinking water, fluoride supplements, infant formula, fluoride toothpaste, and other factors. Neurath (32) recently published an analysis of the NHANES 1999–2004 and 2011–2012 fluorosis data and examined several of these factors. Additional research into these and other factors and their contributions to dental fluorosis prevalence in the United States may still be needed.

Summary

There are potential sources of error in the measurement of fluorosis in general, and specifically in sample surveys such as NHANES. For time-varying factors, these may include intra-subject (within subject) variation and intra- or inter-examiner (within or between examiner) variations

Figure 7. Cumulative distribution of dental fluorosis severity levels, based on person-level Dean's Fluorosis Index values, for youth aged 6–19 years, by age group: National Health and Nutrition Examination Survey, 1999–2004 and 2011–2016



SOURCE: NCHS, National Health and Nutrition Examination Survey, 1999–2004 and 2011–2016.

in measurement. Within-subject variation could occur if there were changes in the underlying disease process between assessments. However, dental fluorosis is the result of exposure to fluoride from when a child is born up to about 8 years of age. Therefore, intra-subject variability due to physiological or biological reasons is unlikely. Intra- or inter-examiner error, however, may occur due to lack of adherence to the protocol by examiners and changes in quality control procedures over time. During 1999–2004 and 2011–2016, there were only minor changes to the dental fluorosis assessment protocol (no dental explorer used for assessment in 2011–2016) and the QA procedures (repeat examinations by the same examiner only occurred in 1999–2001). There was no change to the measurement scale (DFI). This index is, however, known for its subjective nature, which leads to potential variability, especially at the low end of the index where distinguishing between very mild (i.e., less than 25% of the tooth) and mild (25% to less than 50%) can be difficult. Evaluation of gold standard examinations found that agreement (across four different weighting schemes) ranged from 0.51 to 0.98 for the nine primary dental examiners in 1999–2016. These values indicate moderate (0.41–0.60) to almost-perfect (0.81–0.99) agreement based on Landis and Koch, and mostly adequate agreement (33 out of 36 kappas ≥ 0.6) based on McHugh. It is possible, however, that the quantity of repeats as a percentage of the total examined persons (2.8%–3.6%) was inadequate for making conclusions on the reliability of a tool with such subjective variability. Additionally, the assessment of intra-examiner reliability for 1999–2001, as published by Dye et al. (2007), did not show perfect agreement in scoring of the same person only a few days later (weighted kappa statistics comparing person-level scores for the same participant by the same examiner ranged from 0.56 to 0.72). This demonstrates the substantial subjectivity and variability of this scoring method. There was no assessment of intra-examiner reliability in 2011–2016, and procedures to ensure calibration of the reference examiner over time are not available. Therefore, the possibility of a shift in how the examiners assessed dental fluorosis over time cannot be ruled out.

The impact of the complex sampling scheme of NHANES and geographic variability in the presence of fluoride in drinking water on the results were also evaluated. While there were significant sample design changes between 1999–2004 and 2011–2016, and specifically with the oversampled groups, when proper weighting procedures are used, the final estimates produced should still be reflective of the civilian noninstitutionalized U.S. population of youth aged 6–19 years, and estimates from the different survey periods should be comparable. There were no changes to the inclusion and exclusion criteria for the examination for youth aged 6–19 years specifically and no differences in the percentage of the sample who completed the fluorosis assessment.

Variability in the prevalence of different dental fluorosis severity levels was seen within each 6-year time period, as well as across the two time periods. For example, during 1999–2004, questionable fluorosis ranged from 0.4% to 34.6% and during 2011–2016, mild fluorosis prevalence ranged from 9.1% to 40.4%. The reasons for this variability cannot be determined but likely include random error, true change in prevalence, or changes in the application of measurement processes. There may be other factors outside of these that could also contribute to the variability in 2-year prevalence estimates.

Further analyses of the synthetic birth cohort born in 1992–1998 revealed that the prevalence of mild or greater fluorosis in the first permanent molars was 9.5% in youth aged 6–9 years in 2001–2004, but was 46.9% in youth aged 16–19 years in 2011–2014. This increase does not seem biologically plausible because fluorosis develops before teeth erupt and, therefore, prevalence in erupted teeth for the same birth cohort should not change over time. The analysis of the synthetic cohort with adjustment for possible changes in the demographic characteristics of the population over time suggests that the observed increase in fluorosis prevalence does not reflect the genuine amount of change within the U.S. population.

Conclusions

NHANES is currently the only survey providing national estimates on dental fluorosis. Two-year estimates of fluorosis prevalence by DFI category demonstrated variability within and across the 6-year time periods. Inter-examiner reliability statistics found that agreement ranged from 0.51 to 0.98 for the nine primary dental examiners in 1999–2016. These values indicate moderate (0.41–0.60) to almost-perfect (0.81–0.99) agreement based on Landis and Koch, and mostly adequate agreement (33 out of 36 kappas ≥ 0.6) based on McHugh. There was also high percent agreement in defining very mild or greater fluorosis. The proportion of the total number of examinations for which a gold standard examination was conducted was, however, low and the DFI scoring method has high examiner subjectivity. The observed increase in dental fluorosis prevalence with age between 2001–2004 and 2011–2014, based on the analysis of the synthetic cohort, is not biologically plausible. This suggests that there may have been some change in the way the examiners evaluated the level of fluorosis over time. The quality assessment findings in this report should be strongly considered when determining whether these data are appropriate for the user's analytic objectives, including studies of prevalence and trends.

References

1. Institute of Medicine. Dietary reference intakes for calcium, phosphorus, magnesium, vitamin D, and fluoride. Washington, DC: National Academies Press. 1997.
2. U.S. Department of Health and Human Services Federal Panel on Community Water Fluoridation. U.S. Public Health Service recommendation for fluoride concentration in drinking water for the prevention of dental caries. Public Health Rep 130(4):318–31. 2015.
3. Koulourides T. Summary of session II: Fluoride and the caries process. J Dent Res 69(2 Suppl):558. 1990.
4. Featherstone JD. Prevention and reversal of dental caries: Role of low level fluoride. Community Dent Oral Epidemiol 27(1):31–40. 1999.
5. Fluoride Recommendations Work Group. Recommendations for using fluoride to prevent and control dental caries in the United States. MMWR Recomm Rep 50(RR–14):1–42. 2001.
6. Dean HT. Classification of mottled enamel diagnosis. J Am Dent Assoc 21(8):1421–6. 1934.
7. Fejerskov O, Manji F, Baelum V. The nature and mechanisms of dental fluorosis in man. J Dent Res 69 Spec No:692–700. 1990.
8. Dean HT. The investigation of psychological effects by the epidemiological method. Fluorine and Dental Health 19:23–33. 1942.
9. Massler M, Schour I. Atlas of the mouth in health and disease. 2nd ed. Chicago, IL: American Dental Association. 1958.
10. Avery JK. Oral development and histology. Baltimore, MD: Williams and Wilkins. 1987.
11. Clark MB, Slayton RL, Section on Oral Health. Fluoride use in caries prevention in the primary care setting. Pediatrics 134(3):626–33. 2014.
12. Kaminsky LS, Mahoney MC, Leach J, Melius J, Miller MJ. Fluoride: Benefits and risks of exposure. Crit Rev Oral Biol Med 1(4):261–81. 1990.
13. National Research Council. Fluoride in drinking water: A scientific review of EPA's standards. Washington, DC: National Academies Press. 2006.
14. Zipf G, Chiappa M, Porter KS, Ostchega Y, Lewis BG, Dostal J. National Health and Nutrition Examination Survey: Plan and operations, 1999–2010. National Center for Health Statistics. Vital Health Stat 1(56). 2013.
15. Curtin LR, Mohadjer LK, Dohrmann SM, Montaquila JM, Kruszon-Moran D, Mirel LB. The National Health and Nutrition Examination Survey: Sample design,

1999–2006. National Center for Health Statistics. Vital Health Stat 2(155). 2012.

16. Johnson CL, Dohrmann SM, Burt VL, Mohadjer LK. National Health and Nutrition Examination Survey: Sample design, 2011–2014. National Center for Health Statistics. Vital Health Stat 2(162). 2014.

17. Dye BA, Barker LK, Selwitz RH, Lewis BG, Wu T, Fryar CD, et al. Overview and quality assurance for the National Health and Nutrition Examination Survey (NHANES) oral health component, 1999–2002. Community Dent Oral Epidemiol 35(2):140–51. 2007.

18. Dye BA, Nowjack-Raymer R, Barker LK, Nunn JH, Steele JG, Tan S, et al. Overview and quality assurance for the oral health component of the National Health and Nutrition Examination Survey (NHANES), 2003–04. J Public Health Dent 68(4):218–26. 2008.

19. Dye BA, Afful J, Thornton-Evans G, Iafolla T. Overview and quality assurance for the oral health component of the National Health and Nutrition Examination Survey (NHANES), 2011–2014. BMC Oral Health. 2019 (in press).

20. Chen TC, Parker JD, Clark J, Shin HC, Rammon JR, Burt VL. National Health and Nutrition Examination Survey: Estimation procedures, 2011–2014. National Center for Health Statistics. Vital Health Stat 2(177). 2018.

21. Johnson CL, Paulose-Ram R, Ogden CL, Carroll M, Kruszon-Moran D, Dohrmann SM, Curtin LR. National Health and Nutrition Examination Survey: Analytic guidelines, 1999–2010. National Center for Health Statistics. Vital Health Stat 2(161). 2013.

22. Russell AL. The differential diagnosis of fluoride and nonfluoride enamel opacities. J Public Health Dent 21(4):143–6. 1961.

23. Landis JR, Koch GG. The measurement of observer agreement for categorical data. Biometrics 33(1): 159–74. 1977.

24. McHugh ML. Interrater reliability: The kappa statistic. Biochem Med (Zagreb) 22(3):276–82. 2012.

25. Cicchetti DV, Allison T. A new procedure for assessing reliability of scoring EEG sleep recordings. American J EEG Technol 11(3):101–10. 1971.

26. Fleiss JL, Cohen J. The equivalence of weighted kappa and the intraclass correlation coefficient as measures of reliability. Educ Psychol Meas 33:613–9. 1973.

27. Kumar JV, Swango PA, Opima PN, Green EL. Dean's fluorosis index: An assessment of examiner reliability. J Public Health Dent 60(1):57–9. 2000.

28. Beltran-Aguilar ED, Barker L, Dye BA. Prevalence and severity of dental fluorosis in the United States, 1999–2004. NCHS Data Brief, no 53. Hyattsville, MD: National Center for Health Statistics. 2010.

29. American Dental Association. For the dental patient. Tooth eruption: The permanent teeth. J Am Dent Assoc 137(1):127. 2006.

30. U.S. Department of Health, Education, and Welfare. Public Health Service: Drinking water standards, revised 1962. Washington, DC: U.S. Public Health Service. 1962.

31. Mascarenhas AK. Risk factors for dental fluorosis: A review of the recent literature. Pediatr Dent 22(4):269–77. 2000.

32. Neurath C, Limeback H, Osmunson B, Connett M, Kanter V, Wells CR. Dental fluorosis trends in US oral health surveys: 1986 to 2012. JDR Clin Trans Res. 2019.

Table 1. Dean's Fluorosis Index criteria and scoring on the National Health and Nutrition Examination Survey dental fluorosis clinical assessment: 1999–2004 and 2011–2016

NHANES DFI value	Fluorosis severity level	Description
0	Normal	No fluorosis detected
1	Very mild	Opaque, paper white areas involving less than 1/4 of the tooth surface
2	Mild	Opaque, paper white areas involving 1/4 to less than 1/2 of the tooth surface
3	Moderate	Opaque paper white areas involving 1/2 or more of the tooth surface
4	Severe	Discrete or confluent pitting in involved areas
5	Questionable	Slight aberration of normal enamel appearance, including white flecks
8	Nonfluoride opacity	Coded if nonfluoride opacity
9	Cannot be assessed	Coded if the tooth was missing, not fully erupted, or 1/2 or more of the tooth was replaced with a restoration, covered with orthodontic band, or destroyed by caries

NOTES: NHANES is National Health and Nutrition Examination Survey. DFI is Dean's Fluorosis Index.

SOURCE: *NHANES Oral Health Examiners Manual*, available from: https://wwwn.cdc.gov/nchs/data/nhanes/2015-2016/manuals/2016_Oral_Health_Examiners_Procedures_Manual.pdf.

Table 2. Examples of Dean's Fluorosis Index values at tooth and person level: National Health and Nutrition Examination Survey dental fluorosis clinical examination: 1999–2004 and 2011–2016

Person	DFI value for 28 assessed teeth																												DFI for person[1]
A	1	1	1	1	0	0	0	0	0	0	1	1	1	1	1	1	1	1	0	0	0	0	0	0	1	1	1	1	1
B	3	2	3	3	3	4	1	1	3	2	3	3	3	2	3	3	2	2	2	1	1	1	1	3	3	2	3	3	3
C	3	3	2	2	1	2	2	2	2	2	2	2	3	3	3	3	3	3	1	1	1	1	1	1	3	3	2	3	3

[1]Person-level DFI value assigned as the lower of the two most affected tooth-level values.

NOTES: DFI is Dean's Fluorosis Index. 0 = Normal, 1 = Very mild, 2 = Mild, 3 = Moderate, 4 = Severe.

SOURCE: NCHS, National Health and Nutrition Examination Survey, 1999–2004 and 2011–2016.

Table 3. Qualitative interpretation of numeric kappa statistic values

Kappa statistic	Interpretation
Less than 0.0	Poor agreement
0.00–0.20	Slight agreement
0.21–0.40	Fair agreement
0.41–0.60	Moderate agreement
0.61–0.80	Substantial agreement
0.81–0.99	Almost perfect

SOURCE: See reference 23 in this report.

Table 4. Number and percentage of survey participants aged 6–19 years who were examined at the National Health and Nutrition Examination Survey mobile examination center and eligible for the dental fluorosis assessment, by age and race and Hispanic origin: 1999–2004 and 2011–2016

Characteristic	1999–2004					2011–2016				
	MEC examined[1]	Eligible for dental fluorosis exam[2]		Complete fluorosis exam[3]		MEC examined[1]	Eligible for dental fluorosis exam[2]		Complete fluorosis exam[3]	
	n	*n*	Percent	*n*	Percent	*n*	*n*	Percent	*n*	Percent
Total	10,235	9,826	96.0	9,395	91.8	7,752	7,601	98.1	7,158	92.3
Age (years)										
6–11	3,255	3,171	97.4	2,774	85.2	3,861	3,791	98.2	3,378	87.5
12–19	6,980	6,655	95.3	6,621	94.9	3,891	3,810	97.9	3,780	97.1
Race and Hispanic origin										
Non-Hispanic white	2,644	2,544	96.2	2,420	91.5	1,957	1,931	98.7	1,811	92.5
Non-Hispanic black	3,258	3,135	96.2	3,019	92.7	2,008	1,965	97.9	1,870	93.1
Mexican American	3,502	3,348	95.6	3,200	91.4	1,643	1,607	97.8	1,531	93.2
Hispanic (including Mexican American)	…	…	…	…	…	2,526	2,475	98.0	2,333	92.4
Non-Hispanic Asian	…	…	…	…	…	798	774	97.0	724	90.7

… Category not applicable.
[1]MEC is mobile examination center. MEC examined column determined by total number of participant records in the Dentition/Fluorosis public-release data files.
[2]Participant who received the Oral Health Exam (OHX) (complete or partial OHX status code).
[3]Participant had at least two teeth with non-missing dental fluorosis values (0, 1, 2, 3, 4, 5). Nonfluoride opacity (code 8) and cannot be assessed (code 9) values were converted to missing.

SOURCE: NCHS, National Health and Nutrition Examination Survey, 1999–2004 and 2011–2016.

Table 5. Different weighting schemes used for computing kappa statistics for comparing person-level Dean's Fluorosis Index values assigned by dental examiner and reference examiner: 1999–2004 and 2011–2016

DFI value	Weighting scheme					
			Unweighted kappa weights			
0	1.000	0.000	0.000	0.000	0.000	0.000
0.5	0.000	1.000	0.000	0.000	0.000	0.000
1	0.000	0.000	1.000	0.000	0.000	0.000
2	0.000	0.000	0.000	1.000	0.000	0.000
3	0.000	0.000	0.000	0.000	1.000	0.000
4	0.000	0.000	0.000	0.000	0.000	1.000
			Cicchetti-Allison kappa weights[1]			
0	1.000	0.875	0.750	0.500	0.250	0.000
0.5	0.875	1.000	0.875	0.625	0.375	0.125
1	0.750	0.875	1.000	0.750	0.500	0.250
2	0.500	0.625	0.750	1.000	0.750	0.500
3	0.250	0.375	0.500	0.750	1.000	0.750
4	0.000	0.125	0.250	0.500	0.750	1.000
			Fleiss-Cohen kappa weights[2]			
0	1.000	0.984	0.938	0.750	0.438	0.000
0.5	0.984	1.000	0.984	0.859	0.609	0.234
1	0.938	0.984	1.000	0.938	0.750	0.438
2	0.750	0.859	0.938	1.000	0.938	0.750
3	0.438	0.609	0.750	0.938	1.000	0.938
4	0.000	0.234	0.438	0.750	0.938	1.000
			Custom #1 kappa weights			
0	1.000	0.667	0.333	0.000	0.000	0.000
0.5	0.667	1.000	0.667	0.333	0.000	0.000
1	0.333	0.667	1.000	0.667	0.333	0.000
2	0.000	0.333	0.667	1.000	0.667	0.333
3	0.000	0.000	0.333	0.667	1.000	0.667
4	0.000	0.000	0.000	0.333	0.667	1.000
			Custom #2 kappa weights			
0	1.000	0.500	0.000	0.000	0.000	0.000
0.5	0.500	1.000	0.500	0.000	0.000	0.000
1	0.000	0.500	1.000	0.500	0.000	0.000
2	0.000	0.000	0.500	1.000	0.500	0.000
3	0.000	0.000	0.000	0.500	1.000	0.500
4	0.000	0.000	0.000	0.000	0.500	1.000

[1]See reference 25 in this report.
[2]See reference 26 in this report.

NOTES: DFI is Dean's Fluorosis Index. 0 = Normal, 0.5 = Questionable (recoded from collected value of 5), 1 = Very mild, 2 = Mild, 3 = Moderate, 4 = Severe.

SOURCE: NCHS, National Health and Nutrition Examination Survey, 1999–2004 and 2011–2016.

Table 6. Percent agreement and kappa statistics on person-level Dean's Fluorosis Index values assigned by dental examiner and reference examiner for participants aged 6–19 years with available gold standard observations, by 2-year survey cycle: National Health and Nutrition Examination Survey, 1999–2004 and 2011–2016

Survey years and examiner	*n*	Agreement (percent)	Unweighted kappa	Weighted Cicchetti-Allison	Weighted Fleiss-Cohen	Weighted Custom #1	Weighted Custom #2
1999–2000							
A	37	67.6	0.57	0.72	0.87	0.69	0.63
B	53	73.6	0.64	0.77	0.87	0.78	0.76
2001–2002							
A	25	44.0	0.31	0.50	0.65	0.48	0.42
C	48	72.9	0.61	0.70	0.81	0.70	0.69
D	27	66.7	0.58	0.73	0.83	0.71	0.66
2003–2004							
D	70	61.4	0.48	0.70	0.88	0.65	0.62
E	38	47.4	0.35	0.62	0.80	0.57	0.51
2011–2012							
F	34	79.4	0.74	0.82	0.88	0.83	0.81
G	26	84.6	0.78	0.93	0.98	0.87	0.84
2013–2014							
F	32	46.9	0.31	0.53	0.71	0.50	0.46
H	37	59.5	0.40	0.57	0.71	0.55	0.53
2015–2016							
H	23	78.3	0.60	0.73	0.85	0.69	0.68
I	25	80.0	0.69	0.81	0.90	0.73	0.71

NOTES: All kappa statistics were computed using the following Dean's Fluorosis Index values: 0 = Normal, 0.5 = Questionable (recoded from collected value of 5), 1 = Very mild, 2 = Mild, 3 = Moderate, 4 = Severe. All values of 8 (nonfluoride opacity) and 9 (cannot be assessed) were set to missing.

SOURCE: NCHS, National Health and Nutrition Examination Survey, 1999–2004 and 2011–2016.

Table 7. Percent agreement and kappa statistics on person-level Dean's Fluorosis Index values assigned by dental examiner and reference examiner for participants aged 6–19 years with available gold standard observations, by 6-year survey period: National Health and Nutrition Examination Survey, 1999–2004 and 2011–2016

Survey years and examiner	*n*	Agreement (Percent)	Unweighted kappa	Weighted Cicchetti-Allison	Weighted Fleiss-Cohen	Weighted Custom #1	Weighted Custom #2
1999–2004							
A	62	58.1	0.45	0.62	0.77	0.60	0.54
B	53	73.6	0.64	0.77	0.87	0.78	0.76
C	48	72.9	0.61	0.70	0.81	0.70	0.69
D	97	62.9	0.51	0.71	0.86	0.68	0.64
E	38	47.4	0.35	0.62	0.80	0.57	0.51
2011–2016							
F	66	63.6	0.53	0.70	0.82	0.69	0.65
G	26	84.6	0.78	0.93	0.98	0.87	0.84
H	60	66.7	0.48	0.64	0.77	0.62	0.60
I	25	80.0	0.69	0.81	0.90	0.73	0.71

NOTES: All kappa statistics were computed using the following Dean's Fluorosis Index values: 0 = Normal, 0.5 = Questionable (recoded from collected value of 5), 1 = Very mild, 2 = Mild, 3 = Moderate, and 4 = Severe. All values of 8 (nonfluoride opacity) and 9 (cannot be assessed) were set to missing.

SOURCE: NCHS, National Health and Nutrition Examination Survey, 1999–2004 and 2011–2016.

Table 8. Disagreement on person-level Dean's Fluorosis Index values assigned by dental examiner and reference examiner for participants aged 6–19 years with available gold standard observations, by 6-year survey period: National Health and Nutrition Examination Survey, 1999–2004 and 2011–2016

Survey years	Number of gold standard exams	Disagree[1]		Examiner value greater than reference value	Reference value greater than examiner value	Differ by 1 level	Differ by 2 levels	Differ by 3 levels	Differ by 4 levels	Differ by 5 levels
		n	Percent	Percent						
1999–2004	339	127	37.5	48.8	51.2	82.7	15.0	2.4	–	–
2011–2016	198	62	31.3	54.8	45.2	88.7	11.3	–	–	–

– Quantity zero.
[1]Disagreements are based on differences in person-level Dean's Fluorosis Index values: 0 = Normal, 0.5 = Questionable (recoded from collected value of 5), 1 = Very mild, 2 = Mild, 3 = Moderate, and 4 = Severe. For example, examiner value = questionable and reference value = very mild is reported as differing by 1 level.

SOURCE: NCHS, National Health and Nutrition Examination Survey, 1999–2004 and 2011–2016.

Table 9. Percentage of survey participants aged 6–19 years with gold standard observations classified at each person-level Dean's Fluorosis Index value: National Health and Nutrition Examination Survey, 1999–2004 and 2011–2016

DFI level	1999–2004 (*n* = 339)		2011–2016 (*n* = 198)	
	Dental examiner	Reference examiner	Dental examiner	Reference examiner
Normal	26.0	30.1	13.1	11.6
Questionable	25.4	20.1	12.6	14.6
Very mild	29.8	26.8	34.9	39.9
Mild	13.0	15.3	23.7	17.7
Moderate	5.0	5.6	13.1	13.1
Severe	0.9	2.1	2.5	3.0

NOTE: DFI is Dean's Fluorosis Index.

SOURCE: NCHS, National Health and Nutrition Examination Survey, 1999–2004 and 2011–2016.

Table 10. Percentage (weighted) dental fluorosis severity levels, based on person-level Dean's Fluorosis Index, among youth aged 6–19 years: National Health and Nutrition Examination Survey, 1999–2004 and 2011–2016

Survey years	*n*	Normal	Questionable	Very mild	Mild	Moderate	Severe
		Percent (SE)					
1999–2004	9,395	44.3 (2.8)	19.2 (1.6)	25.3 (1.5)	7.7 (0.5)	3.2 (0.4)	0.4 (0.1)
1999–2000	3,103	65.9 (4.6)	0.4 (0.2)	22.3 (3.1)	7.0 (0.9)	3.9 (1.1)	0.5 (0.2)
2001–2002	3,326	48.8 (5.4)	20.6 (3.6)	21.8 (2.1)	6.4 (0.8)	2.1 (0.3)	0.2 (0.1)
2003–2004	2,966	20.0 (2.8)	34.6 (1.8)	31.9 (2.6)	9.5 (1.0)	3.5 (0.7)	0.5 (0.2)
2011–2016	7,158	18.1 (1.6)	10.4 (1.1)	35.6 (2.5)	21.5 (2.2)	13.4 (1.5)	1.0 (0.2)
2011–2012	2,304	34.1 (4.0)	8.8 (0.9)	19.5 (2.1)	14.9 (1.6)	20.6 (3.7)	2.0 (0.5)
2013–2014	2,502	6.1 (1.2)	5.7 (1.3)	28.7 (4.2)	40.4 (5.2)	18.4 (2.1)	0.8 (0.3)
2015–2016	2,352	14.3 (2.9)	16.8 (3.0)	58.4 (4.9)	9.1 (2.2)	1.3 (0.4)	0.1 (0.1)

NOTE: SE is standard error.

SOURCE: NCHS, National Health and Nutrition Examination Survey, 1999–2004 and 2011–2016.

Vital and Health Statistics Series Descriptions

Active Series

Series 1. Programs and Collection Procedures
Reports describe the programs and data systems of the National Center for Health Statistics, and the data collection and survey methods used. Series 1 reports also include definitions, survey design, estimation, and other material necessary for understanding and analyzing the data.

Series 2. Data Evaluation and Methods Research
Reports present new statistical methodology including experimental tests of new survey methods, studies of vital and health statistics collection methods, new analytical techniques, objective evaluations of reliability of collected data, and contributions to statistical theory. Reports also include comparison of U.S. methodology with those of other countries.

Series 3. Analytical and Epidemiological Studies
Reports present data analyses, epidemiological studies, and descriptive statistics based on national surveys and data systems. As of 2015, Series 3 includes reports that would have previously been published in Series 5, 10–15, and 20–23.

Discontinued Series

Series 4. Documents and Committee Reports
Reports contain findings of major committees concerned with vital and health statistics and documents. The last Series 4 report was published in 2002; these are now included in Series 2 or another appropriate series.

Series 5. International Vital and Health Statistics Reports
Reports present analytical and descriptive comparisons of U.S. vital and health statistics with those of other countries. The last Series 5 report was published in 2003; these are now included in Series 3 or another appropriate series.

Series 6. Cognition and Survey Measurement
Reports use methods of cognitive science to design, evaluate, and test survey instruments. The last Series 6 report was published in 1999; these are now included in Series 2.

Series 10. Data From the National Health Interview Survey
Reports present statistics on illness; accidental injuries; disability; use of hospital, medical, dental, and other services; and other health-related topics. As of 2015, these are included in Series 3.

Series 11. Data From the National Health Examination Survey, the National Health and Nutrition Examination Surveys, and the Hispanic Health and Nutrition Examination Survey
Reports present 1) estimates of the medically defined prevalence of specific diseases in the United States and the distribution of the population with respect to physical, physiological, and psychological characteristics and 2) analysis of relationships among the various measurements. As of 2015, these are included in Series 3.

Series 12. Data From the Institutionalized Population Surveys
The last Series 12 report was published in 1974; these reports were included in Series 13, and as of 2015 are in Series 3.

Series 13. Data From the National Health Care Survey
Reports present statistics on health resources and use of health care resources based on data collected from health care providers and provider records. As of 2015, these reports are included in Series 3.

Series 14. Data on Health Resources: Manpower and Facilities
The last Series 14 report was published in 1989; these reports were included in Series 13, and are now included in Series 3.

Series 15. Data From Special Surveys
Reports contain statistics on health and health-related topics from surveys that are not a part of the continuing data systems of the National Center for Health Statistics. The last Series 15 report was published in 2002; these reports are now included in Series 3.

Series 16. Compilations of Advance Data From Vital and Health Statistics
The last Series 16 report was published in 1996. All reports are available online; compilations are no longer needed.

Series 20. Data on Mortality
Reports include analyses by cause of death and demographic variables, and geographic and trend analyses. The last Series 20 report was published in 2007; these reports are now included in Series 3.

Series 21. Data on Natality, Marriage, and Divorce
Reports include analyses by health and demographic variables, and geographic and trend analyses. The last Series 21 report was published in 2006; these reports are now included in Series 3.

Series 22. Data From the National Mortality and Natality Surveys
The last Series 22 report was published in 1973. Reports from sample surveys of vital records were included in Series 20 or 21, and are now included in Series 3.

Series 23. Data From the National Survey of Family Growth
Reports contain statistics on factors that affect birth rates, factors affecting the formation and dissolution of families, and behavior related to the risk of HIV and other sexually transmitted diseases. The last Series 23 report was published in 2011; these reports are now included in Series 3.

Series 24. Compilations of Data on Natality, Mortality, Marriage, and Divorce
The last Series 24 report was published in 1996. All reports are available online; compilations are no longer needed.

For answers to questions about this report or for a list of reports published in these series, contact:

Information Dissemination Staff
National Center for Health Statistics
Centers for Disease Control and Prevention
3311 Toledo Road, Room 4551, MS P08
Hyattsville, MD 20782

Tel: 1–800–CDC–INFO (1–800–232–4636)
TTY: 1–888–232–6348
Internet: https://www.cdc.gov/nchs
Online request form: https://www.cdc.gov/info
For e-mail updates on NCHS publication releases, subscribe online at: https://www.cdc.gov/nchs/govdelivery.htm.

U.S. DEPARTMENT OF
HEALTH & HUMAN SERVICES

Centers for Disease Control and Prevention
National Center for Health Statistics
3311 Toledo Road, Room 4551, MS P08
Hyattsville, MD 20782–2064

For more NCHS Series Reports, visit:
https://www.cdc.gov/nchs/products/series.htm

DHHS Publication No. (PHS) 2019–1383, Series 2, No. 183
CS303088