

## Clustering Tooth Surfaces into Biologically Informative Caries Outcomes

J.R. Shaffer, E. Feingold, X. Wang, D.E. Weeks, R.J. Weyant, R. Crout, D.W. McNeil and M.L. Marazita  
*J DENT RES* 2013 92: 32 originally published online 11 October 2012  
DOI: 10.1177/0022034512463241

The online version of this article can be found at:  
<http://jdr.sagepub.com/content/92/1/32>

---

Published by:



<http://www.sagepublications.com>

On behalf of:

[International and American Associations for Dental Research](#)

**Additional services and information for *Journal of Dental Research* can be found at:**

**Email Alerts:** <http://jdr.sagepub.com/cgi/alerts>

**Subscriptions:** <http://jdr.sagepub.com/subscriptions>

**Reprints:** <http://www.sagepub.com/journalsReprints.nav>

**Permissions:** <http://www.sagepub.com/journalsPermissions.nav>

>> [Version of Record](#) - Dec 10, 2012

[OnlineFirst Version of Record](#) - Oct 11, 2012

[What is This?](#)

J.R. Shaffer<sup>1,2\*</sup>, E. Feingold<sup>1-3</sup>,  
X. Wang<sup>2,4,5</sup>, D.E. Weeks<sup>1,3</sup>,  
R.J. Weyant<sup>2,6</sup>, R. Crout<sup>2,7</sup>,  
D.W. McNeil<sup>8</sup>, and M.L. Marazita<sup>1,2,4,5,9</sup>

<sup>1</sup>Department of Human Genetics, Graduate School of Public Health, University of Pittsburgh, Pittsburgh, PA 15261, USA; <sup>2</sup>Center for Oral Health Research in Appalachia, University of Pittsburgh, Pittsburgh, PA 15261, USA, and West Virginia University, Morgantown, WV 26506, USA; <sup>3</sup>Department of Biostatistics, Graduate School of Public Health, University of Pittsburgh, Pittsburgh, PA 15261, USA; <sup>4</sup>Center for Craniofacial and Dental Genetics, School of Dental Medicine, University of Pittsburgh, Pittsburgh, PA 15219, USA; <sup>5</sup>Department of Oral Biology, School of Dental Medicine, University of Pittsburgh, Pittsburgh, PA 15261, USA; <sup>6</sup>Department of Dental Public Health and Information Management, University of Pittsburgh, School of Dental Medicine, Pittsburgh, PA 15261, USA; <sup>7</sup>Department of Periodontics, West Virginia University School of Dentistry, Morgantown, WV 26506, USA; <sup>8</sup>Dental Practice and Rural Health, West Virginia University School of Dentistry, Morgantown, WV 26506, USA; and <sup>9</sup>Clinical and Translational Science Institute, and Department of Psychiatry, School of Medicine, University of Pittsburgh, Pittsburgh, PA, USA; \*corresponding author, jrs51@pitt.edu

*J Dent Res* 92(1):32-37, 2013

## ABSTRACT

Dental caries affects most adults worldwide; however, the risk factors for dental caries do not necessarily exert their effects uniformly across all tooth surfaces. Instead, the actions of some risk factors may be limited to a subset of teeth/surfaces. Therefore, we used hierarchical clustering on tooth surface-level caries data for 1,068 Appalachian adults (ages 18-75 yrs) to group surfaces based on co-occurrence of caries. Our cluster analysis yielded evidence of 5 distinct groups of tooth surfaces that differ with respect to caries: (C1) pit and fissure molar surfaces, (C2) mandibular anterior surfaces, (C3) posterior non-pit and fissure surfaces, (C4) maxillary anterior surfaces, and (C5) mid-dentition surfaces. These clusters were replicated in a national dataset (NHANES 1999-2000, N = 3,123). We created new caries outcomes defined as the number of carious tooth surfaces within each cluster. We show that some cluster-based caries outcomes are heritable (*i.e.*, under genetic regulation;  $p < 0.05$ ), whereas others are not. Likewise, we demonstrate the association between some cluster-based caries outcomes and potential risk factors such as age, sex, educational attainment, and toothbrushing habits. Together, these results suggest that the permanent dentition can be subdivided into groups of tooth surfaces that are useful for understanding the factors influencing cariogenesis. **Abbreviations:** COHRA, Center for Oral Health in Appalachia, the principal study sample; C1-5, clusters 1-5, groups of similarly behaving tooth surfaces identified through hierarchical clustering; DMFS index, decayed, missing, or filled surfaces, a traditional caries measure representing the number of affected surfaces across the entire dentition; DMFS1-5, partial DMFS indices representing the number of affected surfaces within a hierarchical cluster; and NHANES, National Health and Nutrition Examination Survey, the secondary study sample.

**KEY WORDS:** dental caries, permanent dentition, white spots, hierarchical clustering, cluster analysis, heritability.

DOI: 10.1177/0022034512463241

Received June 29, 2012; Last revision August 30, 2012; Accepted September 8, 2012

A supplemental appendix to this article is published electronically only at <http://jdr.sagepub.com/supplemental>.

© International & American Associations for Dental Research

# Clustering Tooth Surfaces into Biologically Informative Caries Outcomes

## INTRODUCTION

Dental caries, which affects the great majority of adolescents and adults throughout the world, is a multi-factorial disease caused by the effects of numerous environmental, behavioral, and genetic factors. Many risk factors have been identified, such as: host genetics (Horowitz *et al.*, 1958); environmental exposures, including fluoride, cariogenic bacteria, and pH-altering agents; behavioral factors, including diet and oral hygiene; characteristics of the dentition, including enamel composition and positions and morphology of teeth; characteristics of the oral environment, including saliva composition, flow rate, and pH buffering capacity; and demographic factors, including age, sex, race, ethnicity, socio-economic status, and access to oral health care (Hunter, 1988). This complexity is further compounded by the disease phenotype, which may manifest as innumerable combinations of caries lesions across tooth surfaces of the full dentition. Because caries risk factors may exert differential effects across tooth surfaces of the permanent dentition, measurable caries experience may be modeled as the cumulative result of multiple superimposed patterns of decay due to the various risk factors (Batchelor and Sheiham, 2004; Shaffer *et al.*, 2012a).

In epidemiological studies, however, caries experience is typically reduced to a single measure of decay, such as the widely used DMFT/S index (calculated as the sum of decayed, missing due to decay, or filled/restored teeth/surfaces). Such global measures of caries experience ignore the fact that categories of tooth surfaces exhibit differences in susceptibility to decay and are differentially affected by risk factors. Because they ignore the patterns of dental caries across the dentition, global measures of decay such as DMFT/S index may be limited in their ability to identify caries risk factors that exert their effects on specific categories of tooth surfaces. Indeed, previous studies have demonstrated that modeling the patterns of tooth decay is beneficial for epidemiological (Psoter *et al.*, 2003, 2009) and genetic studies (Shaffer *et al.*, 2012a, 2012b). We used hierarchical clustering analysis to group tooth surfaces into categories based on co-occurrence of caries lesions. We then generated novel caries outcomes reflecting these tooth-surface categories, and explored their utility for studying caries etiology.

## METHODS

### COHRA Sample Recruitment and Data Collection

The principal sample used in this study was ascertained through the Center for Oral Health Research in Appalachia (COHRA), a joint effort established in 2000 between the University of Pittsburgh and West Virginia University (Polk *et al.*, 2008). COHRA was designed to study the individual-, familial-, and community-based factors affecting oral health in the Appalachian population, an underserved, low-income group with disproportionately poorer oral health than the general US population (Purnell and Counts, 1998). Participants were recruited according to a household-based protocol requiring at least one biological parent-offspring pair and extending participation to all other household members regardless of legal or biological relationships. In total, 732 households were recruited, which comprised 2,663 participants. Written informed consent was provided by adult participants for themselves and for their participating children. All procedures were approved by the Institutional Review Boards of the relevant Universities. Full details on sample recruitment have previously been described (Polk *et al.*, 2008).

Evidence of dental caries experience was visually assessed with a dental explorer for each surface of each tooth during intra-oral examinations conducted by trained and calibrated dentists or research dental hygienists. Inter- and intra-examiner concordances of caries assessments were high (Polk *et al.*, 2008; Wendell *et al.*, 2010). Tooth surfaces were scored as sound, non-cavitated carious, cavitated carious, restored, missing due to caries, or missing due to other reasons. Caries assessment was performed in accordance with the National Center for Health Statistics Dental Examiners Procedures Manual (Section 4.9.1.3) to maximize comparability with other national datasets. This method yields high-quality, reproducible data for research purposes.

Data on several covariates were also collected. Questionnaire and/or interview was used to assess self-reported data on sex, race, age, birth year, education, toothbrushing frequency, and home water source (city/public *vs.* well). Three-minute timed expectoration was used to assess unstimulated saliva flow rate (mL/min). A fluoride-specific electrode was used to measure home water fluoride (mg/L) in water samples.

The present study, which aimed to model the patterns of caries lesions across surfaces of the permanent dentition, was limited to the subset of participants (N = 1,068) aged 18 to 75 yrs. We re-coded each tooth surface (of 128 surfaces, *i.e.*, buccal, distal, lingual, and mesial surfaces on all 28 permanent teeth, plus occlusal surfaces on pre-molars and molars, excluding third molars) as 0 for sound or missing due to reasons other than decay, or 1 for non-cavitated, cavitated, missing due to decay, or restored. Note that previous work on this sample has shown that modeling non-cavitated lesions is beneficial for genetics studies (Wang *et al.*, 2010). Applying this coding scheme yielded a matrix of 1,068 participants by 128 surface-level caries affection statuses, which was used as input for cluster analysis.

### Hierarchical Clustering

We used Ward's minimum variance method (Ward, 1963), a type of agglomerative hierarchical clustering, to group similarly

affected tooth surfaces (*i.e.*, surfaces that exhibit the same caries affection status), starting with each surface as its own cluster, and then successively merging similar clusters to form a clustering hierarchy. The Euclidean distance (*i.e.*, a common measure of statistical similarity between 2 vectors) was used to determine similarity among clusters. Clustering was performed with the R statistical environment (R Foundation for Statistical Computing, Vienna, Austria).

### Statistical Analysis

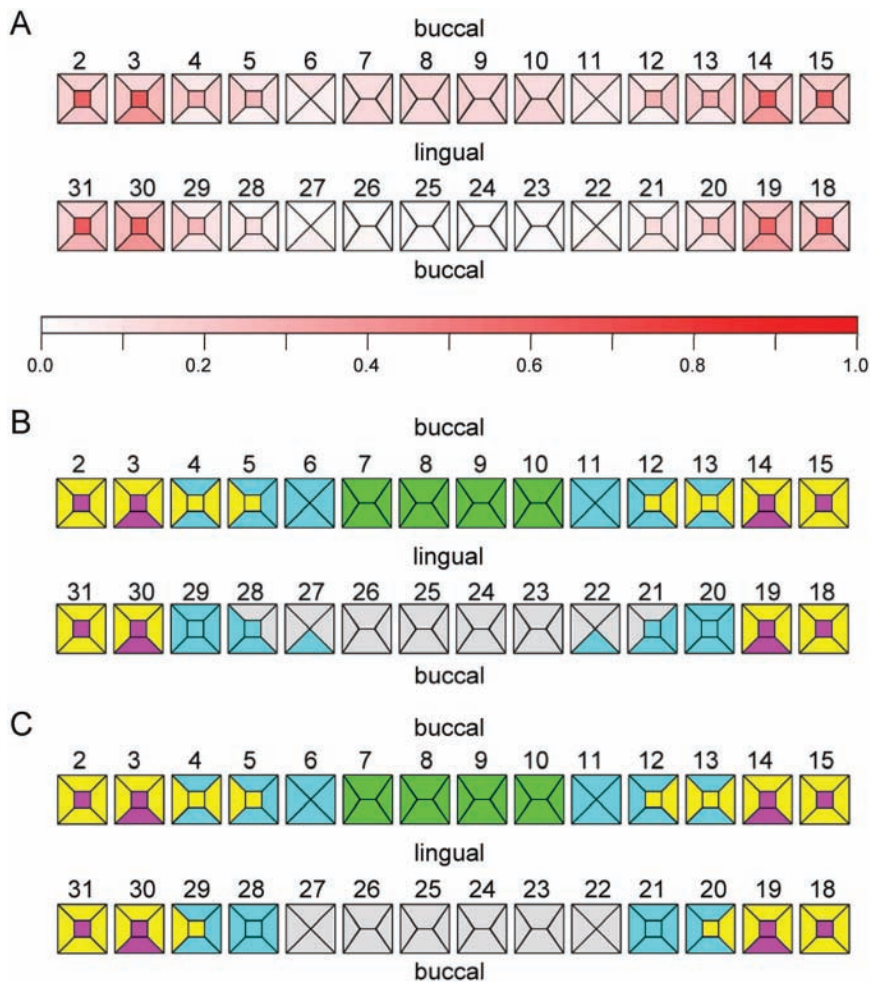
The data matrix of 1,068 participants by 128 tooth-surface-specific caries affection statuses was used as input for hierarchical clustering, which generated clusters of similar tooth surfaces (with respect to occurrence of caries lesions). To help determine stability of the clusters, we repeated the hierarchical clustering on ten randomly chosen subsets of 50% of the full sample. Likewise, because our sample contained related individuals, we repeated our hierarchical clustering using the maximal set of unrelated individuals. Overall, clusters were very robust (see Appendix).

To create novel dental caries outcomes, we calculated the partial DMFS index (*i.e.*, the number of decayed, missing due to decay, or filled surfaces) restricted to surfaces included in each cluster. To test the association of these novel caries outcomes with several potential risk factors, we calculated non-parametric Spearman correlations. Sex, race, home water source (which is related to urban *vs.* rural residency), and toothbrushing frequency were treated as binary categorical variables, with race dichotomized as white *vs.* non-white, and toothbrushing frequency dichotomized as 1 or less *vs.* 2 or more brushings *per* day. Age, age<sup>2</sup>, birth year (included to assess cohort effects), saliva flow, and home water source fluoride were treated as continuous variables. Educational attainment was coded as 0 to 2, respectively, for up to high school diploma, some college, and undergraduate or advanced degree. For all association tests, unadjusted p values (shown) and family-adjusted p values [calculated with SOLAR (Almasy and Blangero, 1998)] were nearly identical.

We used variance decomposition methods as implemented in SOLAR (Almasy and Blangero, 1998) to calculate the heritability of caries outcomes while adjusting for age and age<sup>2</sup> (which were the strongest risk factors observed in the correlational analysis). This method models phenotypic variance among all types of relatives as a function of the theoretical degree of genetic sharing (*i.e.*, that parents and offspring share 50% of their genome, siblings share 50%, half-siblings share 25%, unrelated individuals share 0%, etc.). Details for this method as applied to the COHRA sample have previously been reported (Wang *et al.*, 2010; Shaffer *et al.*, 2012a, 2012b). The heritability estimate is interpreted as the proportion of phenotypic variance attributable to the cumulative effect of all genes.

### NHANES Data

For comparison with our COHRA sample, we also performed hierarchical clustering on the dental caries data available for the NHANES 1999-2000 cohort. Analysis was limited to the subset of 3,123 participants aged 18 to 75 yrs for whom dental caries assessments were made. Caries data were recoded to form a



**Figure 1.** Clustering tooth surfaces based on co-occurrence of caries. **(A)** Surface-level caries prevalence. Each tooth surface is delineated and shaded. Top row represents maxillary teeth (2-15). Bottom row represents mandibular teeth (18-31). **(B)** Cluster membership by surface in COHRA and **(C)** NHANES: (magenta) cluster 1 includes pit and fissure molar surfaces; (gray) cluster 2 includes mandibular anterior surfaces; (yellow) cluster 3 includes posterior non-pit and fissure surfaces; (green) cluster 4 includes maxillary anterior surfaces; and (cyan) cluster 5 includes mid-dentition surfaces. Clusters were generated by hierarchical clustering of tooth-surface-level caries data without any information on the relative positions of tooth surfaces across the permanent dentition.

matrix of 3,123 participants by 128 surfaces, where 0 represents no observed evidence of caries and 1 represents caries or restoration. Third molars were excluded.

## RESULTS

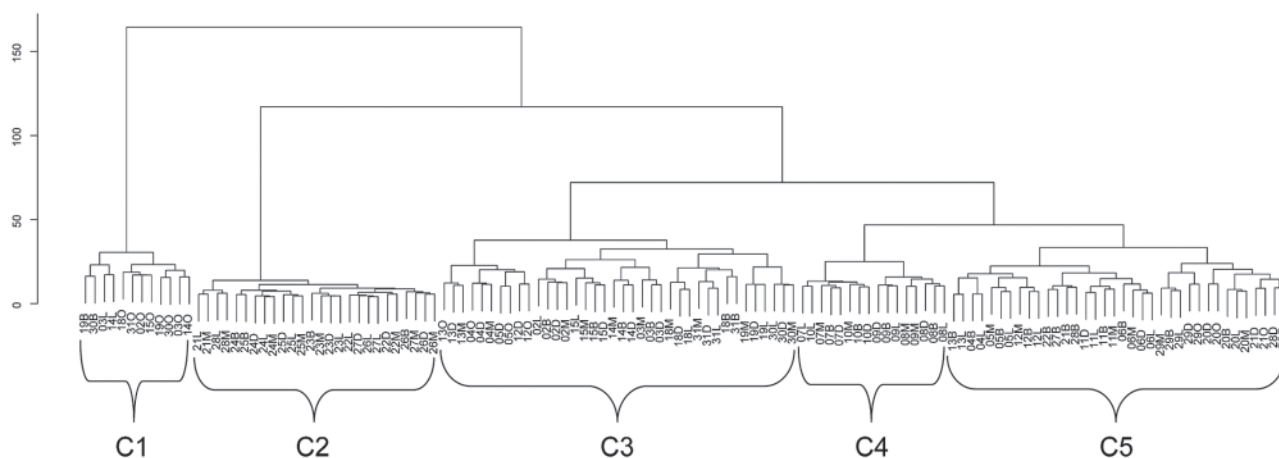
Descriptive characteristics of the COHRA study population are shown in the Appendix Table. The sample ( $N = 1,068$ ) was 63% female and 90% white, with a mean age of 34.7 yrs. Educational attainment was low (16% college graduates, 25% some college). Prevalence of dental caries for each tooth surface, shown in Fig. 1A, was higher in COHRA than in the general US population.

Fig. 2 shows the dendrogram representing the hierarchical clustering of tooth surfaces in COHRA. We identified 5 clusters (C1-C5) of tooth surfaces based on similarity in occurrence of

dental caries: (C1) pit and fissure molar surfaces, (C2) mandibular anterior surfaces, (C3) posterior non-pit and fissure surfaces, (C4) maxillary anterior surfaces, and (C5) the mid-dentition surfaces (Fig. 1B). It is important to note that we generated these surface clusters from caries data only, without any information regarding their relative anatomic positions in the permanent dentition. Interestingly, cluster analysis generated contiguous and symmetrical clusters that sensibly recreated the spatial and morphological relationships among tooth surfaces. Cross-validation showed that clusters were robust (see Appendix). Moreover, clusters generated from COHRA were nearly identical to clusters generated from the 1999-2000 NHANES data (Fig. 1C). For both the full COHRA sample and NHANES sample, cluster 5 was further partitioned into sub-categories comprised of maxillary vs. mandibular teeth, although this separation was not consistent across random sub-samples comprised of 50% of the COHRA sample (results not shown).

To create novel dental caries outcomes, we calculated for each participant the number of carious surfaces within each of the 5 clusters, which we denote as DMFS1, DMFS2, DMFS3, DMFS4, and DMFS5. We tested the univariate associations of DMFS1-DMFS5 with a number of potential caries risk factors (Table 1). All caries outcomes were strongly associated with age, age<sup>2</sup>, and birth year ( $p < 0.001$ ). Other risk factors were associated only with specific clusters, such as the associations of education with DMFS1, DMFS2, and DMFS4 ( $p$  values  $< 0.001$ ), the suggestive association of sex with DMFS4 and DMFS5 ( $p$  values  $< 0.05$ ), and the suggestive associations of DMFS4 with toothbrushing frequency, race, and home water source ( $p$  values  $< 0.05$ ). Several associations (such as those with education) were observed for specific clusters that were not observed for the global DMFS index. Many significant covariates had weak effect sizes, which is consistent with the multi-factorial nature of dental caries.

Based on trait similarity among biological relatives in our sample, we estimated the heritability ( $h^2$ , *i.e.*, the proportion of trait variation due to the cumulative effect of all genetic factors) of DMFS1-DMFS5 (Table 2). DMFS2 ( $h^2 = 54\%$ ,  $p = 0.002$ ), DMFS3 ( $h^2 = 43\%$ ,  $p = 0.004$ ), and DMFS5 ( $h^2 = 40\%$ ,  $p = 0.008$ ) were significantly heritable, whereas DMFS1 was suggestively heritable ( $h^2 = 27\%$ ,  $p = 0.06$ ), and DMFS4 was not heritable ( $h^2 = 0\%$ ,  $p = 0.5$ ).



**Figure 2.** Dendrogram showing the cluster hierarchy and 5 outcome clusters (C1-C5) in COHRA. Individual tooth surfaces are annotated by tooth number (02-15, 18-31) and orientation (B, D, L, M, O, respectively, for buccal, distal, lingual, mesial, and occlusal). Y-axis is the cluster criterion, Ward's minimum within-cluster variance. Each individual tooth surface is indicated by a node at the bottom of the Fig. Moving up the Fig., each node is connected to its closest neighbor node (representing either an individual tooth surface or a cluster of multiple surfaces) by branches proportional to the similarity of caries experience between nodes.

**Table 1.** Univariate (Spearman) Correlations between Caries Outcomes and Possible Risk Factors

Variable	DMFS		DMFS1		DMFS2		DMFS3		DMFS4		DMFS5	
	cor	p value	cor	p value	cor	p value	cor	p value	cor	p value	cor	p value
Sex	0.04	0.17	0.03	0.32	-0.02	0.52	0.04	0.16	<b>-0.07</b>	<b>0.03</b>	<b>0.06</b>	<b>0.05</b>
Race (white vs. other)	0.01	0.79	-0.03	0.29	0.02	0.42	-0.04	0.21	<b>0.08</b>	<b>0.01</b>	-0.02	0.52
Age (yrs)	<b>0.19</b>	<b>4E-10</b>	<b>0.16</b>	<b>2E-7</b>	<b>0.12</b>	<b>4E-5</b>	<b>0.27</b>	<b>5E-19</b>	<b>0.15</b>	<b>1E-6</b>	<b>0.25</b>	<b>3E-17</b>
Age <sup>2</sup> (yrs)	<b>0.19</b>	<b>4E-10</b>	<b>0.16</b>	<b>2E-7</b>	<b>0.12</b>	<b>4E-5</b>	<b>0.27</b>	<b>5E-19</b>	<b>0.15</b>	<b>1E-6</b>	<b>0.25</b>	<b>3E-17</b>
Birth yr	<b>-0.15</b>	<b>6E-7</b>	<b>-0.12</b>	<b>8E-5</b>	<b>-0.11</b>	<b>2E-4</b>	<b>-0.24</b>	<b>8E-15</b>	<b>-0.14</b>	<b>4E-6</b>	<b>-0.24</b>	<b>7E-15</b>
Educational attainment	0.01	0.76	<b>0.13</b>	<b>3E-5</b>	<b>-0.08</b>	<b>8E-3</b>	0.04	0.17	<b>-0.08</b>	<b>6E-3</b>	-0.02	0.54
Saliva flow	-0.03	0.33	0.01	0.69	-0.01	0.72	-0.02	0.46	0.01	0.83	0.00	1.00
Home water source fluoride	0.06	0.16	0.04	0.41	<b>0.09</b>	<b>0.03</b>	0.08	0.06	0.04	0.41	0.07	0.12
Toothbrushing frequency	-0.02	0.44	0.00	0.93	-0.02	0.53	0.00	0.92	<b>-0.08</b>	<b>0.02</b>	-0.01	0.72
Water source (public vs. well)	0.00	0.94	0.03	0.28	0.00	0.90	0.04	0.24	<b>-0.06</b>	<b>0.05</b>	-0.03	0.43

Bold = associations with p values ≤ 0.05.  
 Bonferroni-adjusted alpha for 6 caries outcomes = 0.008.  
 Bonferroni-adjusted alpha for 10 risk factors = 0.005.  
 Bonferroni-adjusted alpha for 60 individual tests = 0.0008.

**DISCUSSION**

In this study, we applied hierarchical clustering to tooth-surface-level caries data to group tooth surfaces based on similarity in dental caries experience. Five clusters were observed in our COHRA sample, which were nearly identical to the results of cluster analysis in the NHANES sample, suggesting that these clusters are reproducible in national datasets, and are not unique to the Appalachian population. Moreover, the 5 clusters were very consistent across randomly partitioned halves of our sample, further indicating their robustness to the inclusion of any specific individuals. The clusters exhibited perfect left-right symmetry, and were comprised of sensible, mostly contiguous, tooth surfaces. Despite the fact that the clustering method did not use any information on

the positions of the teeth, 4 of the 5 clusters reflected the positional relationships of surfaces across the permanent dentition: anterior vs. posterior and maxillary vs. mandibular (whereas one cluster was comprised of highest-risk pit and fissure molar surfaces).

We generated cluster-specific caries outcomes (DMFS1-DMFS5; representing the number of carious surfaces in a cluster) and showed that some were heritable, indicating the causal role of genetic factors on caries experience, whereas others were not. In particular, DMFS2 (comprised of the anterior mandibular surfaces) was the most heritable outcome, which corresponds to the group of surfaces with the lowest prevalences of caries. In contrast, DMFS4 (comprised of the maxillary incisors), was not heritable (*i.e.*,  $h^2 = 0\%$ ), and corresponds to surfaces with fairly high prevalence of caries. Interestingly, in young children, caries

**Table 2.** Heritability of Caries Outcomes Adjusted for Age and Age<sup>2</sup>

Phenotype	h <sup>2</sup>	SE	p value	R <sup>2</sup>
Traditional DMFS index	0.42	0.16	0.007	0.05
DMFS1	0.27	0.17	0.057	0.03
DMFS2	0.54	0.18	0.003	0.04
DMFS3	0.43	0.15	0.004	0.10
DMFS4	0.00	-	0.500	0.03
DMFS5	0.40	0.16	0.008	0.08
Maxillary <sup>a</sup>	0.34	0.19	0.041	0.07
Mandibular	0.31	0.14	0.016	0.06

h<sup>2</sup> = heritability (*i.e.*, the proportion of trait variation attributable to the cumulative role of genes).

SE = standard error of the heritability estimate.

R<sup>2</sup> = proportion of trait variation attributable to age and age<sup>2</sup>; unadjusted models were similar (results not shown).

<sup>a</sup>DMFS5 may be subdivided into maxillary and mandibular components.

of the maxillary incisors is associated with nighttime bottle feeding (Douglass *et al.*, 2001) and access to sweets (Johnsen *et al.*, 1984); we speculate that environmental or behavioral factors may similarly affect maxillary incisors in adults, perhaps overshadowing any genetic predispositions. This reasoning is supported by the observed associations of DMFS4 with toothbrushing frequency and home water source. The high heritability of some caries outcomes indicates that they may be better suited for gene mapping than the traditional DMFS index (h<sup>2</sup> = 42%). Indeed, in another study we have used the cluster-based caries outcomes in a genome-wide association study (Shaffer *et al.*, 2013). Likewise, results showing that different clusters were associated with different potential risk factors, including associations that were not observed for the traditional DMFS index, suggest that clusters may serve as useful outcomes for traditional epidemiological studies, in addition to genetics studies. (See the Appendix for an expanded discussion of these results.)

One issue warranting consideration is how to decide on the number of distinct clusters without over-fitting. Various criteria can be imposed on the clustering hierarchy to delineate clusters—that is, the level of within-cluster similarity and between-cluster dissimilarity that defines the clusters. Rather than choosing an *a priori* distance threshold for determining this, we used two-fold cross-validation in which we repeated hierarchical clustering in randomly chosen halves of the sample (see Appendix). This process demonstrated 5 stable clusters. However, in the full COHRA sample, the next would-be clusters were subdivisions of DMFS5 into maxillary and mandibular surfaces; these subdivisions were also observed for NHANES. Interestingly, heritability of DMFS5 as a whole was greater than each of these subdivisions.

In addition to cross-validation, this study benefits from several other strengths, including our large sample of related individuals with surface-level caries assessments, which facilitated hierarchical clustering and heritability estimation. Also, we were able to replicate our clusters in an independent national dataset. Despite these strengths, however, some limitations merit discussion. First, the method of caries assessment by visual inspection (in both COHRA and NHANES), while appropriate for research purposes, may under-represent caries prevalence. Additionally, teeth missing due to caries (for which all surfaces are counted as carious) and two-surface restorations (*i.e.*, whereby approximal lesions are

accessed through the occlusal surface) may have resulted in the erroneous coding of truly non-carious surfaces adjacent to carious surfaces. The effect of such assessment errors would bias the clustering analysis in favor of including multiple surfaces of the same tooth in the same cluster. However, given that 4 of the 5 observed clusters reflected the positional relationships of surfaces in the permanent dentition (*e.g.*, anterior vs. posterior, and maxillary vs. mandibular), this potential bias was likely not detrimental to the clustering.

In conclusion, this study demonstrates the utility of clustering methods for grouping tooth surfaces with similar caries experience, and reinforces the complexity of dental caries etiology. Refining caries phenotype definitions based on the type of surface may assist in identifying the environmental and genetic risk factors that manifest as specific decay patterns, ultimately leading to improved understanding of the multifactorial nature of dental caries. This study is one of few but much-needed efforts to develop novel, biological data-driven outcomes for investigating dental caries.

## ACKNOWLEDGMENTS

Funding support for this work was provided by the National Institute of Dental and Craniofacial Research, including grants R01-DE014899 and R03-DE021425. Additional support was provided by the University of Pittsburgh School of Dental Medicine, the West Virginia University School of Dentistry, and Eberly College of Arts and Sciences. The content presented herein is solely the responsibility of the authors and does not necessarily represent the official views of the National Institute of Dental and Craniofacial Research or the National Institutes of Health. The funding sources had no input in study design, data collection, analysis, decision to publish, or manuscript preparation. (See the Appendix for full acknowledgments.) The authors declare no potential conflicts of interest with respect to the authorship and/or publication of this article.

## REFERENCES

Almasy L, Blangero J (1998). Multipoint quantitative-trait linkage analysis in general pedigrees. *Am J Hum Genet* 62:1198-1211.

- Batchelor PA, Sheiham A (2004). Grouping of tooth surfaces by susceptibility to caries: a study in 5-16 year-old children. *BMC Oral Health* 4:2.
- Douglass JM, Tinanoff N, Tang JM, Altman DS (2001). Dental caries patterns and oral health behaviors in Arizona infants and toddlers. *Community Dent Oral Epidemiol* 29:14-22.
- Horowitz SL, Osborne RH, DeGeorge FV (1958). Caries experience in twins. *Science* 128:300-301.
- Hunter PB (1988). Risk factors in dental caries. *Int Dent J* 38:211-217.
- Johnsen DC, Schultz DW, Schubot DB, Easley MW (1984). Caries patterns in Head Start children in a fluoridated community. *J Public Health Dent* 44:61-66.
- Polk DE, Weyant RJ, Crout RJ, McNeil DW, Tarter RE, Thomas JG, et al. (2008). Study protocol of the Center for Oral Health Research in Appalachia (COHRA) etiology study. *BMC Oral Health* 8:18.
- Psoter WJ, Zhang H, Pendrys DG, Morse DE, Mayne ST (2003). Classification of dental caries patterns in the primary dentition: a multidimensional scaling analysis. *Community Dent Oral Epidemiol* 31:231-238.
- Psoter WJ, Pendrys DG, Morse DE, Zhang HP, Mayne ST (2009). Caries patterns in the primary dentition: cluster analysis of a sample of 5169 Arizona children 5-59 months of age. *Int J Oral Sci* 1:189-195.
- Purnell LD, Counts M (1998). Appalachians. In: *Transcultural health care: a culturally competent approach*. Purnell LD, editor. Philadelphia, PA: FA Davis.
- Shaffer JR, Feingold E, Wang X, Tcuenco K, Weeks DE, DeSensi RS, et al. (2012a). Heritable patterns of tooth decay in the permanent dentition: principal components and factor analyses. *BMC Oral Health* 12:7.
- Shaffer JR, Wang X, DeSensi RS, Wendell S, Weyant RJ, Cuenco KT, et al. (2012b). Genetic susceptibility to dental caries on pit and fissure and smooth surfaces. *Caries Res* 46:38-46.
- Shaffer JR, Feingold E, Wang X, Lee M, Cuenco KT, Weeks DE, et al. (2013). GWAS of dental caries patterns in the permanent dentition. *J Dent Res* 92:38-44.
- Wang X, Shaffer JR, Weyant RJ, Cuenco KT, DeSensi RS, Crout R, et al. (2010). Genes and their effects on dental caries may differ between primary and permanent dentitions. *Caries Res* 44:277-284.
- Ward JH (1963). Hierarchical grouping to optimize an objective function. *J Am Statist Assoc* 48:236-244.
- Wendell S, Wang X, Brown M, Cooper ME, DeSensi RS, Weyant RJ, et al. (2010). Taste genes associated with dental caries. *J Dent Res* 89:1198-1202.