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

[1]Department of Human Genetics, Graduate School of Public Health, University of Pittsburgh, Pittsburgh, PA 15261, USA; [2]Center for Oral Health Research in Appalachia, University of Pittsburgh, Pittsburgh, PA 15261, USA, and West Virginia University, Morgantown, WV 26506, USA; [3]Department of Biostatistics, Graduate School of Public Health, University of Pittsburgh, Pittsburgh, PA 15261, USA; [4]Center for Craniofacial and Dental Genetics, School of Dental Medicine, University of Pittsburgh, Pittsburgh, PA 15219, USA; [5]Department of Oral Biology, School of Dental Medicine, University of Pittsburgh, Pittsburgh, PA 15261, USA; [6]Department of Dental Public Health and Information Management, University of Pittsburgh, School of Dental Medicine, Pittsburgh, PA 15261, USA; [7]Department of Periodontics, West Virginia University School of Dentistry, Morgantown, WV 26506, USA; [8]Dental Practice and Rural Health, West Virginia University School of Dentistry, Morgantown, WV 26506, USA; and [9]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 DOI: 10.1177/0022034512463241

Clustering Tooth Surfaces into Biologically Informative Caries Outcomes

APPENDIX

COHRA Population Characteristics

Descriptive characteristics of the COHRA sample are shown in the Appendix Table. Compared with the general US population, this Appalachian sample from rural West Virginia and Pennsylvania is poorer, less educated, and underserved with respect to oral health care.

Cross-validation and Defining Separate Clusters

Determining the number of separate clusters is the principal challenge in the field of cluster analysis, and there is currently very little statistical theory in this area. Defining *a priori* minimum distances between clusters is arbitrary and may lead to over-fitting. Therefore, we instead performed two-fold cross-validation (*e.g.,* Salvador and Chan, 2004), a commonly used approach for determining the number of clusters. Cross-validation allows us to determine the sensitivity of our clusters to perturbations of the input data and to identify what level of within-cluster similarity and between-cluster dissimilarity defines stable clusters. We randomly divided our sample into two halves and performed hierarchical clustering on each half. This process was repeated for 10 random halves. By comparing dendrograms, we determined the maximum number of separate clusters that were consistently observed across all random halves. Overall, tooth surfaces were distributed across 5 very stable clusters. Example cluster results for two complementary halves (which total to the full sample) are shown in Appendix Fig. 1. Similarly, to assess any effects of the inclusion of biological relatives in our sample, we repeated hierarchical clustering in the maximal subset of unrelated individuals (see Appendix Fig. 2), which were nearly identical to the total COHRA and NHANES 1999-2000 samples.

In both the total COHRA sample and in the NHANES 1999-2000 sample, the fifth cluster was subdivided into maxillary and mandibular components (indicating possibly 6 instead of 5 clusters). However, these subdivisions were not consistently observed *via* cross-validation of the COHRA sample. That is, in some randomly chosen halves of the COHRA sample, the sixth cluster did not distinguish maxillary and mandibular components of the fifth cluster. Therefore, we have conservatively shown results for the 5 stable clusters, as well as the maxillary and mandibular subdivisions of the fifth cluster.

Overwhelmingly, the cluster results were stable within the COHRA sample, and consistent between the COHRA sample and NHANES 1999-2000 sample. That being stated, there were subtle differences observed among cluster results. For example, left-right asymmetry was observed, albeit rarely, in the random halves (*e.g.,* Appendix Fig. 1A, tooth #20 and tooth #29). Likewise, some tooth surfaces physically positioned on the border between two adjacent clusters shifted membership (*e.g.,* Appendix Figs. 1A and 2, tooth #21 and tooth #28). In many cases, the result of such shifts in cluster membership echoed the subtle differences between COHRA and NHANES 1999-2000. We speculate that the all-or-nothing nature of our clustering approach represents an oversimplification of the relationships among tooth surfaces with respect to dental caries risk factors

Appendix Table. Descriptive Characteristics of the COHRA Sample

Variable	N	% or mean (SD) [range]
Sex (% female)	1,068	63.2
Race: white/black/other	964/73/35	89.9/6.8/3.3
Hispanic (%)	1,052	1.1
Age (yrs)	1,068	34.7 (9.2)
Birth yr	1,066	1971 [1929-1993]
Body mass index (kg/m^2)	874	29.7 (7.7)
Educational attainment		
Up to high school (%)	611	58.8
Some college (%)	265	25.5
4-year degree or higher (%)	163	15.7
Saliva flow rate (mL/min)	1,030	0.68 (0.48)
Home water fluoride level (mg/L)	536	0.68 (0.42)
Public water source (%)	999	79.3
Toothbrushing twice or more *per* day (%)	983	58.2

and that some surfaces may be better modeled by distributing their membership across multiple clusters. Thorough exploration of other methods of cluster analysis may provide added insight into this issue, but is outside the scope of the current manuscript.

Comparison of Caries Patterns with Those from Previous Studies

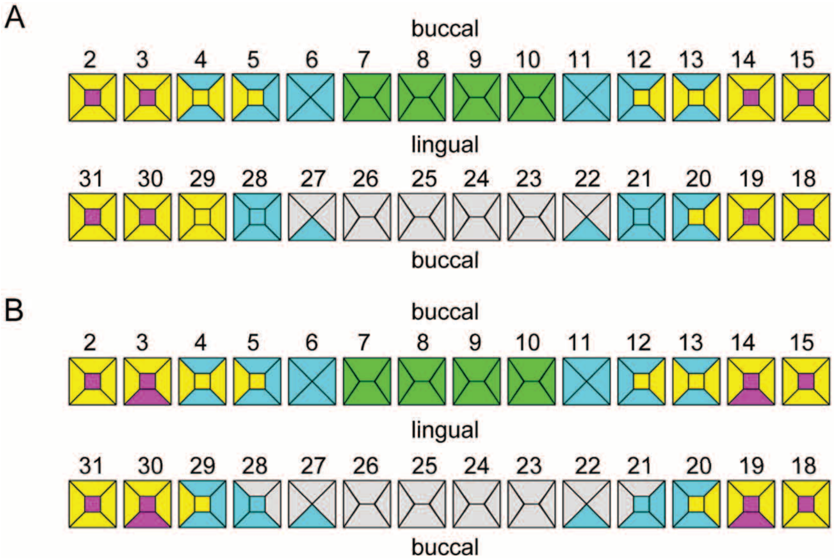
We identified and replicated 5 clusters of tooth surfaces that behave similarly with respect to caries experience. In general, for the anterior teeth (*e.g.*, incisors and maxillary canines), all surfaces of a given tooth tended to be members of the same cluster, whereas for the posterior teeth (*e.g.*, molars and premolars) and mandibular canines, surfaces of a tooth were divided among two different clusters. This observation is reasonable given the comparatively similar and disparate morphologies of surfaces on anterior and posterior teeth, respectively. Our 5 clusters were: (cluster 1) pit and fissure molar surfaces, (cluster 2) mandibular anterior surfaces, (cluster 3) posterior non-pit and fissure surfaces, (cluster 4) maxillary anterior surfaces, and (cluster 5) the mid-dentition surfaces.

Because previous studies of caries patterns have usually assumed *a priori* classification of tooth surfaces, which often differed among studies, detailed comparisons across studies is difficult. However, some commonalities across studies were observed. For example, Douglass *et al.* (1994, 1995) investigated 4 *a priori* patterns of decay in children, 2 of which closely matched the agnostic patterns observed in our study: The “fissure” and “maxillary anterior” patterns were similar to clusters 1 and 4, respectively, in the present study. The “posterior proximal” and “posterior buccal/lingual” patterns studied by Douglass *et al.* did not individually match any of our clusters, though, together, included surfaces from clusters 3 and 5. However, Douglass *et al.* studied the primary dentition, whereas we studied the permanent dentition, so differences in posterior decay patterns may be expected, especially given the inclusion of 8 permanent premolars in our study, which lack analogous teeth in the primary dentition. Another study by Douglass *et al.* (2001) looked at 3 decay patterns in children, “pit and fissure”, “maxillary anterior”, and “posterior proximal”. Again, the “pit and fissure” and “maxillary anterior” patterns were similar to our clusters 1 and 4.

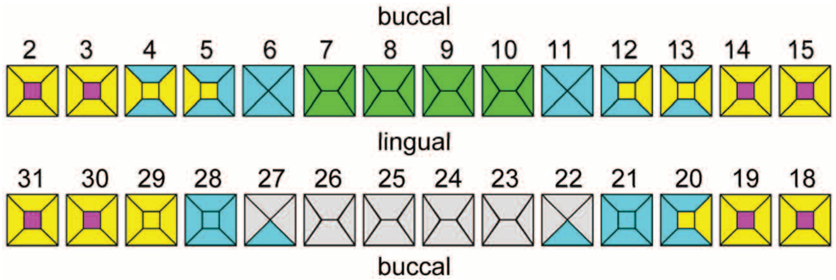
A study by Greenwell *et al.* considered 3 *a priori* caries patterns in children which they called “faciolingual”, “pit and fissure”, and “molar-approximal” (1990). Their “pit and fissure” pattern was similar to our cluster 1, but their other patterns were unlike any of our clusters. The underlying premise of these studies, like ours, was that different risk factors may lead to different patterns of caries. However, in contrast to *a priori* caries patterns, which may be informed by clinical experience but which are not supported by empirical data, our agnostically determined clusters may better represent the differential effects of risk factors.

Other studies have defined caries patterns based on categorizing participants into different classes of dental caries experience. For example, in Johnsen *et al.* (1986), participants were categorized as caries-free or exactly 1 of 5 *a priori* caries patterns, 2 of which were also defined by enamel defects. Given the differences in approach—that is, categorizing participants as opposed to clustering tooth surfaces—their *a priori* caries patterns are difficult to compare with our clusters. However, they showed that water fluoridation was associated with smooth-surface patterns of decay, but not with other patterns of decay. This result is consistent with our study, which showed modest effects of home water source fluoride concentration on cluster 2, the anterior mandibular surfaces, but not on other surfaces.

An innovative study by Psoter *et al.* (2003) used multidimensional scaling to collapse the information across 5,169 children (*i.e.,* 5,169 original dimensions) to map each tooth surface onto 2 new orthogonal dimensions. This multidimensional scaling approach is nearly identical to the principal components analysis previously performed in our sample (Shaffer *et al.,* 2012); however, we performed dimension reduction across tooth surfaces to create new pattern-based caries phenotypes, whereas Psoter *et al.* performed dimension reduction across participants to understand the relationships among surfaces with respect to



Appendix Figure 1. Cluster membership by surface in random halves (**A** and **B**, which sum to the total sample) of the COHRA sample: (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.



Appendix Figure 2. Cluster membership by surface in the maximal set of unrelated individuals from the COHRA sample: (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.

caries experience. Interestingly, both of the new dimensions created in Psoter *et al.* were related to surface morphology and timing of eruption. Given that timing of tooth eruption is correlated with position in the mouth (*i.e.*, anterior *vs*. posterior, and mandibular *vs.* maxillary), their results mesh nicely with our observation that clusters were related to surface morphology and position. Their agnostic procedure, which assumed only contralateral symmetry, suggested 4 caries patterns: (1) maxillary incisor surfaces, (2) first molar occlusal surfaces, (3) second molar pit and fissure surfaces, and (4) all other surfaces. Only their maxillary incisors pattern matched our results (*i.e.*, our cluster 4). However, like other previous studies, Psoter *et al.* looked at primary dentition. A follow-up study in this sample used a variety of cluster analysis methods to group tooth surfaces (Psoter *et al*., 2009), very much like the method used in our study. Their follow-up analyses largely confirmed their original caries patterns, but with possible subdivisions representing maxillary *vs*. mandibular surfaces of first molar and second molar patterns. The potential subdivision of Psoter *et al.* clusters into maxillary *vs*. mandibular surfaces echoes what we observed for cluster 5 in our study.

Overall, caries patterns differed among the limited number of previous studies investigating patterns of dental decay, which is not surprising given that most of these presumed *a priori* patterns (Johnsen *et al*., 1986; Greenwell *et al.,* 1990; Douglass *et al.,* 1994, 1995, 2001), and were conducted in children (Johnsen *et al.,* 1986; Greenwell *et al.,* 1990; Douglass *et al.,* 1994, 1995, 2001; Psoter *et al.*, 2003, 2009). However, when these studies are taken together, some consensus patterns have emerged, including a pattern reflecting maxillary anterior surfaces (like cluster 4 in our study) and one or more patterns reflecting pit and fissure surfaces (like cluster 1 in our study). More work is needed to verify which patterns are most useful for investigating and understanding cariogenesis, though, collectively, the body of literature on this topic suggests that these patterns may differ between primary and permanent dentition, and may differ based on the risk factor profiles of the populations under study.

Cluster Analysis *vs.* DMFS and Other Pattern Extraction Methods

Our results showed that global caries experience in the permanent dentition could be partitioned into patterns defined by similarly behaving clusters of tooth surfaces. Some cluster-based phenotypes were more heritable than traditional DMFS index, whereas others were not. Similarly, the cluster-based phenotypes were associated with different suites of environmental risk factors, such as sex, educational attainment, toothbrushing frequency, race, and home water source. Several of these associations were observed for cluster-based phenotypes, but not for the traditional DMFS index. Overall, these results, which demonstrate the potential benefit of modeling caries patterns, are consistent with previous work with principal components and factor analyses used for the extraction of heritable caries patterns (Shaffer *et al*., 2012). In particular, the second strongest caries pattern identified by factor analysis was predominantly due to maxillary incisors, which is very similar to cluster 4 identified in the present study. A benefit of the cluster analysis approach is that the interpretation of results can be clearly mapped back to individual tooth surfaces, whereas for principal components and factor analyses, interpretation of results with respect to actual tooth surfaces is problematic, which greatly diminishes the utility of the latter methods.

ADDITIONAL ACKNOWLEDGMENTS

We are deeply grateful for the contributions of many individuals and organizations whose support was instrumental in the success of this work. In particular, we thank all participating individuals, families, and communities for their involvement in this study. Additionally, we acknowledge the following health care institutions and organizations: The University of Pittsburgh, Bradford, Center for Rural Health Practice, Bradford, PA; McKean County Dental Center, Bradford, PA; Cornerstone Care Community Medical and Dental Center, Burgettstown, PA; UPMC Braddock Hospital, Braddock, PA; Camden-on-Gauley Medical Center, WV; Community Health Clinic of Nicholas County, WV; Richwood Area Community Hospital, WV; Summersville Memorial Hospital, WV; and the Webster County Memorial Hospital, WV; the GORGE Connection Rural Health Education Partnership Board, the Webster-Nicholas Rural Health Education Consortium Board, the West Virginia Rural Health Education Partnerships program, the Nicholas and Webster Boards of Education, and the UPMC Braddock Community Advisory Board.

AUTHORS' CONTRIBUTIONS

JRS conceived and designed this study. RJW, RC, DWM, and MLM conceived and designed the COHRA initiative. JRS analyzed the data and wrote the manuscript. JRS, EF, WX, DEW, RJW, DWM, and MLM interpreted the results and read, revised, and approved the manuscript.

APPENDIX REFERENCES

Douglass JM, Yi W, Zhang BX, Tinanoff N (1994). Dental caries in preschool Beijing and Connecticut children as described by a new caries analysis system. *Community Dent Oral Epidemiol* 22:94-99.

Douglass JM, Yi W, Zhang BX, Tinanoff N (1995). Caries prevalence and patterns in 3-6-year-old Beijing children. *Community Dent Oral Epidemiol* 23:340-343.

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.

Greenwell AL, Johnsen D, DiSantis TA, Gerstenmaier J, Limbert N (1990). Longitudinal evaluation of caries patterns from the primary to the mixed dentition. *Pediatr Dent* 12:278-282.

Johnsen DC, Bhat M, Kim MT, Hagman FT, Allee LM, Creedom RL, *et al.* (1986). Caries levels and patterns in Head Start children in fluoridated and non-fluoridated, urban and non-urban sites in Ohio, USA. *Community Dent Oral Epidemiol* 14:206-210.

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.

Salvador S, Chan P (2004). Determining the number of clusters/segments in hierarchical clustering/segmentation algorithms. *Proc 16th IEEE Intl Conf on Tools with AI* 576-584.

Shaffer JR, Feingold E, Wang X, Tcuenco K, Weeks DE, DeSensi RS, *et al*. (2012). Heritable patterns of tooth decay in the permanent dentition: principal components and factor analyses. *BMC Oral Health* 12:7.