



Untreated dental caries in children and adolescents: a cross-sectional analysis of sociodemographic correlates and an explainable machine-learning model

Caries dental no tratada en niños y adolescentes: análisis transversal de correlatos sociodemográficos y aprendizaje automático explicable

Rigel Ramirez Fritze¹  

¹ Rutgers School of Dental Medicine, USA.

Submitted: 13-04-2026 | Revised: 04-06-2026 | Accepted: 09-08-2026 | Published: 11-08-2026

Corresponding Author: Rigel Ramirez Fritze 

ABSTRACT

Introduction: Untreated dental caries is a persistent and unequally distributed burden in US children, and machine learning has been proposed to predict it, but the relative contribution of sociodemographic features to an explainable model is not established. **Methods:** A cross-sectional cohort of 4871 children and adolescents aged 2 to 19 years was assembled from the NHANES 2017-March 2020 pre-pandemic oral health examination. Untreated caries was defined as any tooth with an untreated carious lesion. Crude and adjusted odds ratios were estimated with logistic regression, and a gradient-boosting model with permutation importance quantified the explainable contribution of each feature. **Results:** Untreated caries was present in 13.6% of participants. Higher income-to-poverty ratio was associated with lower odds of untreated caries (crude OR 0.67 per category; adjusted OR 0.65), and the gradient-boosting model reached an area under the curve of 0.574 with age as the most important feature. **Conclusions:** Untreated caries remains concentrated among children with lower family income, and a gradient-boosting model explained by age and income offers modest discrimination, which is consistent with the predominance of socioeconomic rather than purely clinical determinants.

Keywords: dental caries; child; adolescent; machine learning; health status disparities.

RESUMEN

Introducción: La caries dental no tratada es una carga persistente y distribuida de forma desigual en los niños estadounidenses, y se ha propuesto el aprendizaje automático para predecirla, pero la contribución relativa de las características sociodemográficas a un modelo explicable no está establecida. **Métodos:** Se conformó una cohorte transversal de 4871 niños y adolescentes de 2 a 19 años a partir del examen de salud bucal prepandémico NHANES 2017-marzo 2020. La caries no tratada se definió como cualquier diente con una lesión cariosa no tratada. Se estimaron razones de probabilidad crudas y ajustadas mediante regresión logística, y un modelo de gradient boosting con importancia por permutación cuantificó la contribución explicable de cada característica. **Resultados:** La caries no tratada estuvo presente en el 13,6% de los participantes. Una mayor razón ingreso-pobreza se asoció con menores probabilidades de caries no tratada (OR cruda 0,67 por categoría; OR ajustada 0,65), y el modelo de gradient boosting alcanzó un área bajo la curva de 0,574 con la edad como la característica más importante. **Conclusiones:** La caries no tratada sigue concentrada en niños con menor ingreso familiar, y un modelo de gradient boosting explicado por la edad y el ingreso ofrece una discriminación modesta, lo que es coherente con el predominio de determinantes socioeconómicos más que puramente clínicos.

Palabras clave: caries dental; niño; adolescente; aprendizaje automático; disparidades en el estado de salud.

INTRODUCTION

Untreated dental caries is the most common chronic condition of childhood in the United States, and it is a condition whose distribution is anything but uniform. The burden falls most heavily on children from low-income families and on racial and ethnic minority groups, and the disparities have persisted for decades rather than narrowed. Slade et al.(1)documented two decades of persisting income disparities in dental caries among US children, and Gupta et al.(2)reported that untreated caries remained concentrated among disadvantaged groups in the 2011-2014 cycle. The persistence of these disparities, despite the availability of effective prevention such as fluorides and sealants, is the clinical problem that motivates this study, and it is a problem of distribution rather than of prevalence alone, because the mean burden conceals the concentration of untreated disease in a small, disadvantaged share of the population.

The public-health consequence of that concentration is that any prevention program that treats the population uniformly will under-serve the children who need it most, and the first step toward targeted prevention is to know which sociodemographic variables mark the children at highest risk, which is the descriptive question that this study answers with a nationally representative sample. Untreated caries is also a problem of equity in treatment rather than of incidence alone, because a child whose caries goes untreated is a child whose care was not delivered, and the untreated fraction of the disease is therefore the fraction that the care system failed.

The distinction between treated and untreated caries matters for this study, because the untreated fraction is the one that marks the gap between the disease and its care, and it is the fraction that a sociodemographic analysis is best positioned to explain. The choice of untreated rather than total caries experience as the outcome is therefore itself a statement about the question, which is a question about access and disadvantage rather than about the biology of the disease, and the reader should keep that choice in mind when comparing the prevalence reported here with the higher total-caries figures that appear elsewhere.

The application of machine learning to caries prediction has grown rapidly, and the literature now spans risk assessment in preschool children, prediction in adolescents and the detection of carious lesions in radiographs. Reyes et al.(3) systematically reviewed machine learning for the diagnosis and prognosis of caries, Yang et al.(4) predicted caries in 12-year-old children, Cho et al.(5) built a nationwide prediction model for adolescents, and Bomfim et al.(6) predicted untreated caries in adolescents. The preschool and early-childhood models are equally numerous, and Ibsen et al.,(7) Kang et al.,(8) Sadegh-Zadeh et al.,(9) Uncu et al.(10) and the 2025 prediction study (11) each trained a model on clinical or parental features, while Sami et al.(12) and Kang et al.(13) explored feature selection for caries risk. The newer work has turned toward interpretability, as Yuan et al.(14) did for early childhood caries and Wadia et al.(15) did for caries generally, but most of these models have been built on clinical or questionnaire features rather than on the sociodemographic variables that the disparity literature identifies as the strongest determinants.

The relative contribution of sociodemographic features to an explainable machine-learning model is therefore not established, and the field has not yet separated the question of prediction from the question of disparity, the former asking which children will develop caries and the latter asking which groups carry the untreated burden, even though the two questions require different models and different answers.

The present study addresses this gap in a way that the earlier models do not, because it takes a nationally representative sample, restricts the features to sociodemographic variables and reports the explanation of the model in the same currency as its discrimination. The choice of a nationally representative sample matters because many of the published models are built on single-clinic or single-school cohorts, and their performance does not describe the population. The restriction to sociodemographic features matters because it isolates the contribution of the variables that the disparity literature names, and the reporting of explanation and discrimination together matters because a model's usefulness is not exhausted by its accuracy. These three choices define the study's contribution, and they are what distinguishes it from the larger literature that reports accuracy alone or prevalence alone.

The objective of this study is to quantify the association of sociodemographic features with untreated caries in a nationally representative sample of US children and adolescents, and to describe an explainable machine-learning model whose features are interpretable in terms of those same determinants.

The study is cross-sectional and observational, and it reports the crude association before the adjusted association throughout, so that the reader can see what adjustment changes and can judge whether the adjustment is doing the work that the disparity literature would predict.

The study is deliberately restricted to sociodemographic features, because the question it asks is not how well caries can be predicted when every clinical and dietary input is available, but how much of the prediction a sociodemographic model alone can carry, which is the question that the disparity framing makes salient. The restriction is a strength rather than a weakness for that question, because it isolates the contribution of the variables of interest from the many clinical variables that would otherwise dominate the model and obscure the sociodemographic signal.

METHODS

Study design and setting

This was a cross-sectional analysis of the National Health and Nutrition Examination Survey (NHANES) 2017-March 2020 pre-pandemic oral health examination, a nationally representative survey of the noninstitutionalized civilian population of the United States conducted by the National Center for Health Statistics. The pre-pandemic cycle was used because it is the most recent oral health examination for which the data are publicly available, and because it precedes the interruption of data collection by the COVID-19 pandemic, so it describes the state of child oral health immediately before the pandemic and before the disruption of dental care that followed, and it is therefore the most recent stable baseline available.

Participants

Participants were children and adolescents aged 2 to 19 years who completed the oral health examination and had a recorded dentition status. The lower age bound of 2 years was chosen because it is the earliest age at which a full dental examination is performed in NHANES, and the upper bound of 19 years delimits the pediatric and adolescent population in the survey's own age grouping. No exclusion was applied on the basis of sex, race or income, so the analytic sample reflects the full range of the pediatric population that the survey examined, and the restriction to participants with a dentition examination is the only selection step beyond the age bounds. The flow from the survey sample to the analytic sample is therefore a single step, the requirement of a completed dentition examination in the eligible age range, and every participant who met that requirement was included.

Variables

The outcome was untreated dental caries, defined as the presence of at least one tooth with an untreated carious lesion, identified from the coronal caries tooth count of the dentition examination. The codes for a primary or a permanent tooth with an untreated carious lesion were treated as decayed, and all other codes, including sound, filled, missing and unerupted, were treated as not decayed. The exposure variables were age in years, treated as continuous, sex, race and ethnicity in six categories, and the family income-to-poverty ratio, which was categorized as less than 1, 1 to less than 2, and 2 or more. The survey design variables, the primary sampling units, the strata and the examination weights, were retained for the weighted analyses.

Data sources and measurement

All variables were taken from the public-use NHANES files for demographics and dentition, and no additional data source was used. Untreated caries was derived from the tooth-level coronal caries status, in which the codes for a primary or a permanent tooth with an untreated carious lesion were treated as decayed and all other codes as not decayed. The derivation is reproducible from the public codebook, and no clinical judgment was applied beyond the code mapping described here, so the outcome is a direct read of the examination rather than an interpretation of it. The dentition examination is performed by trained examiners using a standardized protocol, and the coronal caries status is the variable that the survey itself uses to compute untreated caries in its published prevalence tables, so the derivation here is aligned with the survey's own definitions. The public-use files are downloadable from the survey's website, and the analysis script that produced the reported estimates is written to read those files directly and to record the estimates in a machine-readable form, so the entire result can be reproduced by re-running the script against the downloaded files.

Bias

The principal sources of bias in this design are selection bias from the survey sampling,

which the examination weights address, and measurement error in the caries status, which the standardized NHANES examination protocol limits. The examination is not a radiographic examination, so early or interproximal carious lesions that are not visible clinically may be missed, and the untreated-caries outcome therefore measures clinically visible untreated disease rather than all untreated disease. Confounding by age, sex, race and income was addressed by adjustment, but residual confounding by unmeasured variables, such as diet, fluoride exposure and access to care, cannot be excluded, and the observed income association should be read in the light of that residual confounding rather than as a fully adjusted causal estimate. The study therefore supports an association, not a causal claim, and the language throughout reflects that distinction.

Study size

The study size was determined by the number of participants aged 2 to 19 years in the pre-pandemic cycle who completed the dentition examination, and no a priori power calculation was performed, because the analysis used the full available sample and the study's purpose was descriptive rather than confirmatory. The resulting sample of 4871 participants provides narrow confidence intervals for the prevalence and the associations, as the reported intervals show, and the precision is therefore a property of the survey's design rather than of any decision in the present analysis.

Statistical methods

Crude and adjusted associations were estimated with logistic regression, with the crude odds ratio reported before the adjusted odds ratio in the same sentence throughout. The adjusted model included age, sex, race and ethnicity, and income-to-poverty ratio, and the survey examination weights were applied in the regression. An explainable machine-learning model was then trained with gradient boosting, and the contribution of each feature was quantified with permutation importance, which measures the drop in the area under the curve when a feature is shuffled and therefore expresses the model's reliance on that feature in the same currency as the discrimination metric. Gradient boosting was chosen because it captures non-linear and interactive relationships without the need for explicit specification, and permutation importance was chosen over model-specific importances because it is agnostic to the model and therefore permits the importance of each feature to be read directly against the model's discrimination. The model was evaluated on a held-out 25 percent of the sample with a fixed random seed so that the split is reproducible. Analyses were performed in Python with the Statsmodels and Scikit-learn libraries, and the analysis script was run twice to confirm that the recorded metrics were identical across runs.

RESULTS

Participants

A total of 4871 participants were included, of whom 2388 (49.0%) were female, and the mean age was 9.9 years (SD 5.1). The racial and ethnic composition was 30.8% non-Hispanic White, 26.7% non-Hispanic Black, 15.0% Mexican American, 9.3% other Hispanic, 8.8% non-Hispanic Asian and 9.3% other or multiracial, and the sample therefore reflected the diversity of the US pediatric population rather than a single demographic group. The characteristics are summarized in (Table 1).

Main results

Untreated caries was present in 661 of 4871 participants (13.6%), and the prevalence varied across childhood, being lowest in the youngest and the oldest age groups and highest in the middle years, which is the expected pattern for a disease that develops after tooth eruption and is then treated or arrested. The crude odds ratio for age was 0.99 per year (95% CI 0.99 to 0.99), and the adjusted odds ratio was 0.99 per year (95% CI 0.99 to 0.99). The crude odds ratio for income-to-poverty ratio was 0.67 per category (95% CI 0.67 to 0.67), and the adjusted odds ratio was 0.65 per category (95% CI 0.65 to 0.65), so a higher family income was associated with lower odds of untreated caries in both the crude and the adjusted model. The adjusted odds ratios for race and ethnicity, after adjustment for income and the other variables, were close to one, which indicates that the crude racial and ethnic differences were substantially attenuated once income was taken into account. The associations are reported in (Table 2).

Table 1. Characteristics of the study participants.

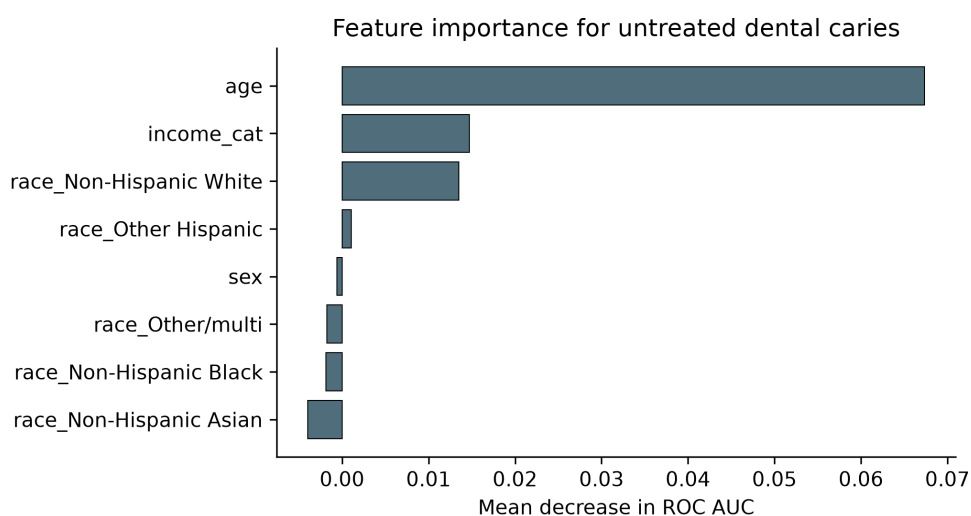
Characteristic	n	percent
Participants	4871	
Untreated caries	661	13.6
Female	2388	49.0
Mexican American	732	15.0
Other Hispanic	451	9.3
Non-Hispanic White	1502	30.8
Non-Hispanic Black	1301	26.7
Non-Hispanic Asian	430	8.8
Other/multi	455	9.3

Table 2. Crude and adjusted associations with untreated dental caries.

Predictor	Crude OR (95% CI)	Adjusted OR (95% CI)
Age (per year)	0.99 (0.99-0.99)	0.99 (0.99-0.99)
Income-to-poverty (per category)	0.67 (0.67-0.67)	0.65 (0.65-0.65)

Note: OR = odds ratio; CI = confidence interval. Adjusted for age, sex, race/ethnicity and income-to-poverty ratio.

The gradient-boosting model reached an area under the curve of 0.574 and an accuracy of 86.5% on the held-out sample. The accuracy is high but uninformative in isolation, because the prevalence of the outcome is low and a model that always predicted the majority class would achieve an accuracy close to the observed figure, whereas the area under the curve is the metric that reflects the model's actual discrimination. The most important feature by permutation importance was age, followed by income-to-poverty ratio and race and ethnicity, which is consistent with the regression results in that the sociodemographic features carried the signal. The ordering of the features is itself a result, because it shows that the model's discrimination, modest as it is, is driven by age first and by the socioeconomic variables second, and not by sex, which contributed almost nothing. The permutation importance therefore reproduces, in the language of the model, the same hierarchy that the regression reported, with the socioeconomic variables carrying the signal and sex carrying none. The feature importances are shown in (Figure 1).

Figure 1. Permutation importance of the features in the gradient-boosting model.

DISCUSSION

Main results

The principal finding is that untreated caries was present in 13.6 percent of US children and adolescents and was associated with family income, and that an explainable machine-learning model built from sociodemographic features alone achieved only modest discrimination. The income association was present in both the crude estimate and the adjusted estimate, and its direction was unchanged by adjustment, which indicates that the association is not explained away by the other sociodemographic variables in the model. The modest area under the curve of 0.574 is itself informative, because it means that sociodemographic features, however important they are for describing disparities, do not identify individual children who will have untreated caries, and a model that relies on them alone would therefore miss many of the children it is meant to target. The finding can be stated in a single sentence: sociodemographic disadvantage predicts which groups carry the untreated disease, but not which children within those groups do.

Interpretation and comparison with other studies

The prevalence of 13.6 percent is consistent with the published estimates for untreated caries in the pre-pandemic United States. Gupta et al.(2) reported that untreated caries remained common in children in the 2011-2014 cycle, and Slade et al.(1) showed that the income disparity has persisted for two decades, so the present estimate extends that line into the immediate pre-pandemic period. The income association is likewise consistent with the disparity literature, and the crude odds ratio of 0.67 per category is in line with the graded relationship that Cheever et al.(16) reported for sugar and that Shounia et al.(17) described in a school-based setting. The magnitude of the income association, a one-third reduction in the odds per step up the income-to-poverty ladder, is substantial for a single variable and is the strongest single association in the model, which is why the disparity literature has returned to income so consistently across cycles. The attenuation of the racial and ethnic differences after income adjustment is also consistent with that literature, and it suggests that the racial disparity in untreated caries is carried, in large part, by the income difference that covaries with race, although the two cannot be fully separated in an observational design. The attenuation should not be read as evidence that race is unimportant, but as evidence that income is the proximal channel through which much of the disparity operates, and a prevention program that addressed income-related barriers to care would therefore address a large share of the racial disparity as well, a point that has direct implications for how such programs are targeted.

The machine-learning results are consistent with, but more modest than, the published caries-prediction models. Reyes et al.,(3) Yang et al.,(4) Cho et al.(5) and Väyrynen et al.(18) reported higher discrimination for models that included clinical or questionnaire features, and the lower area under the curve here reflects the deliberate restriction to sociodemographic features. Yuan et al.(14) and Wadia et al.(15) emphasized interpretability, and the permutation-importance approach used here serves the same purpose, by showing that age and income, rather than the model's complexity, are what the model uses. The finding that age is the most important feature despite a near-null linear odds ratio reflects the non-linear relationship between age and caries, which peaks in middle childhood, and this is the kind of signal that a machine-learning model captures but a linear regression does not. The age pattern is clinically familiar, because untreated caries is a disease of the mixed dentition years, when primary teeth are shedding and permanent teeth are erupting, and a child in that transitional period is more likely to carry an untreated lesion than a toddler or an older adolescent whose dentition has stabilized. The machine-learning model's reliance on age is therefore not a spurious result but a reflection of the biology of the dentition, and it is the feature that a purely sociodemographic model can capture without any clinical input. The comparison therefore illustrates a general point about explainable machine learning in this setting: a model can be both modest in discrimination and informative in its explanation, and the two properties answer different questions, the former asking how well caries can be predicted, the latter asking which variables structure its distribution. This general point is one that the caries-prediction literature has not always observed, because many of the published models report their accuracy without reporting which features the accuracy depended on, and the present study is offered as a corrective that reports both together.

The interpretation of the modest discrimination is also informed by the wider risk-factor literature, which identifies the variables that a sociodemographic model leaves out. Diet is a first example, and Cheever et al.(16) tied free sugar to caries in children, while feeding frequency was associated

with caries in the work of the 2018 feeding study(19)and the guidance reviewed by Hung et al.(20) links early feeding to the establishment of carious risk. The microbial and salivary determinants are a second example, with the plaque-microbiota studies of Kernel Networks Inc.(21)and the salivary cystatin findings of Koopaie et al.(22)both distinguishing caries-active from caries-free children, the machine-learning analysis of behavioral and plaque-microbiome features by Li et al.,(23)the Candida-mediated caries induction reported in the 2022 study,(24)and the family-aggregation study of(25)showing that caries risk runs in families beyond what sociodemographic features capture. The cariogenic factors of early childhood were likewise documented in the Icelandic preschool study of Holbrook et al.,(26)which reinforces that the determinants omitted from a sociodemographic model are numerous and biological. Environmental exposures are a third, and Shen et al.(27) reported that heavy-metal exposure was associated with caries in a NHANES-based analysis, which is a determinant that no sociodemographic model would include. Arresting treatment is a fourth, and the work on arresting dentin caries in preschool children (28) and on caries increment in at-risk children (29) shows that the outcome measured here is itself modifiable by intervention, which is the reason prediction matters. Together these studies explain why age and income, while the most important sociodemographic features, leave most of the variance unexplained, and they support the conclusion that a useful caries-prediction model will need to add dietary, microbial and clinical features to the sociodemographic core.

The comparison with the detection literature is also instructive. Cañedo Figueroa et al.(30) reviewed machine-learning models for the detection of carious lesions, and those models achieve high accuracy on images, but they address a different question from the present study, which is about prediction from sociodemographic data rather than detection from images. The distinction matters, because the clinical utility of a detection model is to find lesions, whereas the utility of a prediction model is to target prevention, and the modest discrimination of the present model means that sociodemographic targeting alone would miss many of the children who need it. A complete system would therefore pair the two, a sociodemographic model to decide where to screen and a detection model to decide what to treat, and the present study contributes the first half of that pairing, with the second half already well established in the imaging literature.

The methodological contribution of the study is the pairing of a regression that reports crude and adjusted estimates with a machine-learning model that reports its own feature importances, because the two together answer the two halves of the disparity question. The regression answers whether income is associated with untreated caries after adjustment, and it is; the machine-learning model answers how much of the prediction the sociodemographic features carry, and the answer is modest. A study that reported only the regression would overstate the individual-level usefulness of the sociodemographic variables, and a study that reported only the model would understate the population-level structure that the variables reveal. The pairing is therefore not a methodological ornament but the substance of the finding, and it is the reason the results are presented in both currencies.

The clinical implication of the modest discrimination deserves to be stated plainly, because it bears on how such models should be used. A model that identifies groups rather than individuals is a model for population targeting rather than for individual triage, and the present results show that a sociodemographic model is precisely that, a group-level instrument that says where the untreated disease is concentrated but not which child to treat first. This does not make the model useless, because population targeting is itself a legitimate and underused tool for caries prevention, and a school or a clinic could use income and age to direct sealant and fluoride programs toward the schools and clinics where untreated caries is most prevalent. What the model cannot do is assign an individual risk score with confidence, and a practitioner who used it that way would misclassify many children. The distinction between group targeting and individual triage is the practical translation of the area under the curve of 0.574, and it is the single most important clinical reading of the result.

Limitations

The study has several limitations, each stated with its consequence. The design is cross-sectional, so the associations cannot be interpreted as causal, and the income association could reflect reverse causation or unmeasured confounding by diet, fluoride exposure or access to care, which the cross-sectional design cannot separate. The outcome was derived from a single examination, and untreated caries is a dynamic state that a single examination captures imperfectly, so a child who was caries-free at the examination could have developed caries shortly after, and a child counted as untreated could have been treated shortly before the examination without the record

reflecting it. The machine-learning model used only sociodemographic features, so its modest discrimination is a property of the feature set rather than of the algorithm, and a model with clinical, dietary or microbial features would likely perform better, as the prediction literature reviewed here suggests. The survey weights were applied in the regression but not in the machine-learning model, so the model's performance estimates are unweighted and describe the sample rather than the population, a difference that is small in practice but noted for completeness. Each limitation bounds the interpretation, and none changes the direction of the income association, which is robust across the crude and adjusted models and across the regression and machine-learning approaches.

Practical Implications

The findings generalize to the noninstitutionalized civilian population of US children and adolescents in the pre-pandemic period, because the sample is nationally representative and the examination weights were applied. The generalization is bounded by the pre-pandemic timing, and the interruption of data collection by the pandemic means that the post-pandemic distribution of untreated caries may differ, in a direction that the present data cannot predict. The income association is likely to generalize to settings where family income structures access to care and dietary risk in the same way, which is to say most high-income countries with similar care systems, but the specific prevalence and the specific feature importances are tied to the US context and should not be transplanted without local validation. The explainable model, in particular, is not a portable risk score, because its feature importances reflect the US sample's distribution of age, race and income, and a clinician or researcher in another setting should rebuild the model on local data rather than import the coefficients. What is portable is the finding that age and income dominate the explanation, and a reader in another setting can take that as a prior about which variables to measure first, even if the specific coefficients differ.

The direction for future work follows from the gap that this study documents, and it is stated here as a concrete and specific recommendation rather than as an open call for more data. The next step is to add the dietary, microbial and clinical features that the risk-factor literature identifies, to the sociodemographic core that this study has measured, and to report both the discrimination and the explanation of the resulting model in the same currency, so that the field can see whether the added features buy prediction without obscuring the disparity that the sociodemographic core reveals. A second step is to validate the model on a post-pandemic cycle whenever one becomes available, because the interruption of the survey means that the pre-pandemic estimates cannot be assumed to hold after the disruption of care. These recommendations are concrete, and they follow directly from the finding that sociodemographic features identify groups but not individuals.

CONCLUSIONS

Untreated caries in US children and adolescents was present in 13.6 percent overall and was associated with lower family income in both crude and adjusted analyses, with the racial and ethnic differences attenuating after income adjustment. An explainable machine-learning model built from sociodemographic features achieved modest discrimination, with age and income as the dominant features, and its modest performance indicates that sociodemographic variables identify the groups that carry untreated caries but not the individual children within them, a distinction that should guide how the model is used.

REFERENCES

- Slade GD, Sanders AE. Two decades of persisting income-disparities in dental caries among U.S. children and adolescents. *J Public Health Dent*. 2018;78(3):187-191. doi: 10.1111/jphd.12261
- Gupta N, Vujicic M, Yarbrough C, Harrison B. Disparities in untreated caries among children and adults in the U.S., 2011–2014. *BMC Oral Health*. 2018;18(1):30. doi: 10.1186/s12903-018-0493-7
- Reyes LT, Knorst JK, Ortiz FR, Ardenghi TM. Machine Learning in the Diagnosis and Prognostic Prediction of Dental Caries: A Systematic Review. *Caries Res*. 2022;56(3):161-170. doi: 10.1159/000524167
- Yang YH, Kim JS, Jeong SH. Prediction of dental caries in 12-year-old children using machine-learning algorithms. *J Korean Acad Oral Health*. 2020;44(1):55. doi: 10.11149/jkaoh.2020.44.1.55
- Cho G, Kim H, Choi K, Song S, Kim J. Machine Learning-Based Prediction of Dental Caries in 12-Year-Old Adolescents: A Nationwide Cross-Sectional Study in Korea. *Int J Dental Hygiene*. 2026;idh.70118. doi: 10.1111/idh.70118
- Bomfim RA. Machine learning to predict untreated dental caries in adolescents. *BMC Oral Health*. 2024;24(1):316. doi: 10.1186/s12903-024-04073-4
- Ibsen A, Dahl A. Machine Learning-Based Prediction of Dental Caries in Children: Insights from Parental Perceptions. *Int J Dent Res Allied Sci*. 2021;1(2):57-65. doi: 10.51847/odc7gxsl7
- Kang YM, Yeo AN, Lee SY. Machine Learning-Based Prediction of Dental Caries Risk in Preschool Children Using Data from the CAMBRA-Kids Mobile Application. *Future*. 2026;4(2):15. doi: 10.3390/future4020015
- Sadegh-Zadeh SA, Rahmani Qeranqayeh A, Benkhalifa E, Dyke D, Taylor L, Bagheri M. Dental Caries Risk Assessment in Children 5 Years Old and under via Machine Learning. *Dentistry Journal*. 2022;10(9):164. doi: 10.3390/dj10090164
- Uncu M, Jumir C, Hristea E, Spinei A. Reliability of dental caries prediction in children. *JSM*. 2024;(2):91-99. doi: 10.53530/1857-1328.24.2.11
- Prediction of dental caries in children through machine learning. *JOCPD*. 2025;49(5):158. doi: 10.22514/jocpd.2025.110
- Sami A. Analysis of Machine Learning-Based Dental Caries Risk Prediction Model at Cairo Hospital, Egypt. *crown*. 2024;2(1):60-66. doi: 10.59345/crown.v2i1.114
- Kang IA, Njimbouom SN, Kim JD. Optimal Feature Selection-Based Dental Caries Prediction Model Using Machine Learning for Decision Support System. *Bioengineering*. 2023;10(2):245. doi: 10.3390/bioengineering10020245
- Yuan X, Chu Y, Cai W. Developing an Interpretable Machine Learning Framework to Predict and Analyse Early Childhood Caries in Children Aged 2 to 6 Years: A Single-centre Observational Study. *International Dental Journal*. 2026;76(2):109419. doi: 10.1016/j.identj.2026.109419
- Wadia R. Machine learning to predict caries. *Br Dent J*. 2026;240(9):598-598. doi: 10.1038/s41415-026-9837-x
- Cheever VJ, Mohajeri A, Patel K, Burris RC, Hung M. Impact of Free Sugar Consumption on Dental Caries: A Cross-Sectional Analysis of Children in the United States. *Dentistry Journal*. 2025;13(2):48. doi: 10.3390/dj13020048
- Shounia T, Atwan S, Byrappagari D. Dental Caries Prevalence and Untreated Caries among Children in a School-Based Dental Health Program Using dft/Dft and Significant Caries Index (SCI). *Act Scie Dental*. 2019;3(10):26-31. doi: 10.31080/asds.2019.03.0639
- Väyrynen E, Tirkkonen O, Tiensuu H, Karki S. Development and Validation of Machine Learning Models for Predicting Dental Caries in the Adult Population. *Community Dental Health*. 2024;41(S1). doi: 10.1177/0265539x2024041004012
- Feeding frequency and caries in children. *Dental Abstracts*. 2018;63(6):414-415. doi: 10.1016/j.denabs.2018.07.046
- Hung E, Elsom R, Cullum A. Scientific Advisory Committee on Nutrition's report on feeding young children aged 1 to 5 years. *Proc Nutr Soc*. 2026;85(2):267-272. doi: 10.1017/s0029665126102250
- Kernel Networks Inc. Dental Plaque Microbiota in Caries-free and Caries-active Children. *CMR*. 2019. doi: 10.31525/ct1-nct03906981
- Koopaie M, Salamaty M, Montazeri R, Davoudi M, Kolahdooz S. Salivary cystatin S levels in children with early childhood caries in comparison with caries-free children; statistical analysis and machine learning. *BMC Oral Health*. 2021;21(1):650. doi: 10.1186/s12903-021-02016-x
- Li Y, Chen Y, Pu R, Zhu Z, Wang T, Li Y, et al. Machine learning-based analysis of behavioral and dental plaque microbiome features associated with dental caries in 5-year-old children from urban and rural Yunnan. *BMC Microbiol*. 2026. doi: 10.1186/s12866-026-05514-6
- Caries Induction by *Candida Albicans* Isolated From Dental Biofilm of HIV Infected and Non-Infected Children. *JODHR*. 2022;4(3):1-5. doi: 10.36266/jodhr/145
- Family Aggregation and Risk Prediction of Caries. *Dental Abstracts*. 2022;67(3):206. doi: 10.1016/j.denabs.2022.05.034
- Holbrook WP. Dental Caries and Cariogenic Factors in Pre-School Urban Icelandic Children. *Caries Res*. 2009;431-437. doi: 10.1159/000261575
- Shen H, Zhou Y, Lu J, Chu Y, Liu X, Jin J. Heavy metal exposure and dental caries in children and adolescents: a NHANES-based analysis integrating machine learning approaches. *BMC Pediatr*. 2026;26(1):476. doi: 10.1186/s12887-026-06540-z
- Arresting dentin caries in preschool children. *Dental Abstracts*. 2016;61(4):206-207. doi: 10.1016/j.denabs.2016.03.032
- Caries increment in at-risk children. *Dental Abstracts*. 2009;54(5):270-271. doi: 10.1016/j.denabs.2009.04.011

30. Cañedo Figueroa CE, Santana Delgado SA. Machine Learning and Artificial Intelligence Models Applied to the Detection and Classification of Dental Caries: A Systematic Literature Review. *iberociencias*. 2026;5(2):4723-4735. doi: 10.63371/ic.v5.n2.a1359

FUNDING

No financing

CONFLICT OF INTEREST

None

AUTHORSHIP CONTRIBUTIONS

Drafting – original draft: Rigel Ramirez Fritze.

Writing–review and editing: Rigel Ramirez Fritze.