Journal of
Oral Health and Oral Epidemiology

https://johoe.kmu.ac.ir
doi:10.34172/.johoe.2509.1793
JOHOE. 2026;15:2509.1793

Original Article

Comparative Performance of Four Prevalence Estimators for Untreated Dental Caries: Application to KERCADRS Phase III study

Amir Hossein Nekouei[1], **Mohammad Reza Baneshi**[2], **Arash Shahravan**[3], **Jahangir Haghani**[4], **Hamid Sharifi**[5], **Aliakbar Haghdoost**[5*]

[1]Physiology Research Center, Institute of Neuropharmacology, Kerman University of Medical Sciences, Kerman, Iran
[2]Australian Women and Girls' Health Research Centre, School of Public Health, Faculty of Medicine, The University of Queensland, Brisbane, QLD, Australia
[3]Endodontology Research Center, Kerman University of Medical Sciences, Kerman, Iran
[4]Oral and Dental Diseases Research Center, Kerman University of Medical Sciences, Kerman, Iran
[5]HIV/STI Surveillance Research Center, and WHO Collaborating Center for HIV Surveillance, Institute for Futures Studies in Health, Kerman University of Medical Sciences, Kerman, Iran

*__Corresponding Author:__ Aliakbar Haghdoost, Email: ahaghdoost@gmail.com

Abstract

Background: The prevalence of untreated dental caries (UDC) is a critical indicator in dental public health. This study evaluated four techniques to estimate UDC prevalence: (1) overall prevalence, (2) average individual prevalence, (3) generalized estimating equations (GEE), and (4) random effects models (REM).

Methods: A simulation study generated hypothetical populations under two scenarios, with intraclass correlation values of 0.05, 0.1, and 0.2: Scenario 1: UDC prevalence (5%, 10%, 20%) independent of missing teeth; Scenario 2: UDC prevalence dependent on the number of missing teeth. Four estimation methods were compared: 1. Overall Prevalence Estimator: calculated as the total number of UDC divided by total teeth; 2. Average Individual Prevalence Estimator: mean of individual prevalence values, 3. GEE: logistic regression with participant-level clustering effects, 4. REM: random effects logistic regression modeling prevalence at both the individual and tooth levels. Performance was assessed using mean squared error (MSE), bias, confidence interval (CI) coverage, and CI length. For practical implications, the simulation study results were applied to estimate the UDC in Phase III of the Kerman Coronary Artery Disease Risk Study (KERCADRS).

Results: When UDC was independent of missing teeth, GEE and the average individual prevalence methods yielded the most reliable estimates (lower MSE and higher CI coverage). When UDC depended on missing teeth, no method performed optimally. However, GEE achieved comparatively better results. Analysis of the KERCADRS data showed a significant correlation between the number of UDC and the number of missing teeth ($r=0.15$, $P<0.001$). Accordingly, the UDC estimated using the GEE method was 26.3% (95% CI: 25.9%, 26.8%).

Conclusion: In contexts where UDC and missing teeth are uncorrelated, GEE and average individual prevalence methods are recommended. When dependencies were present, the GEE method performed slightly better than the other methods. UDC prevalence in Kerman is high, and urgent action is needed to address it.

Keywords: Cluster analysis, Dental caries, Prevalence

Citation: Nekouei AH, Baneshi MR, Shahravan A, Haghani J, Sharifi H, Haghdoost A. Comparative performance of four prevalence estimators for untreated dental caries: application to KERCADRS phase III study. J Oral Health Oral Epidemiology 2026;15:2509.1793. doi:10.34172/.johoe.2509.1793

Received: September 27, 2025, **Revised:** December 8, 2025, **Accepted:** Januray 31, 2026, **ePublished:** May 10, 2026

Introduction

The prevalence of untreated dental caries (UDC) is a critical indicator of dental public health,[1,2] representing the percentage of UDC within a population. This metric is essential for guiding public health policies and programs to control and prevent dental caries,[3,4] a condition that imposes significant financial burdens on health systems and insurers.[5–7] Additionally, accurate prevalence data are necessary for identifying high-risk groups that require targeted interventions.[8] Therefore, precisely estimating UDC prevalence is vital for effective public health planning. However, various methods have been employed to calculate and report this prevalence, each with some limitations.

The most commonly used method is the DMFT index (decayed, missing, and filled teeth), which estimates the average number of teeth affected by caries per individual in a population.[9–11] Although widely used, the DMFT index does not indicate the percentage of UDC in a population, even when averaging component D. Instead, it reflects the cumulative number of teeth affected by decay over an individual's lifetime. Alternatively, some studies report UDC prevalence as the percentage of individuals with at least one untreated decayed tooth.[12–14] However, this approach does not account for the percentage of untreated decayed teeth. Other studies have directly calculated and reported the percentage of UDC.[15–17] Although this method provides useful prevalence data, it poses challenges in calculating the standard deviation.

The complexity of calculating the standard deviation arises from the clustering of teeth within the same mouth. Teeth in an individual's mouth share a similar risk of decay, which varies between individuals. In clustered data, such as teeth, the similarity of an individual's teeth within a cluster is measured using an intraclass correlation. Intraclass correlation, even if small, is important, and ignoring this clustering effect leads to an underestimation of the standard deviation and results in inaccurate statistical inferences.[18–20]

To address these challenges, statistical methods have been developed to account for tooth clustering. Advanced approaches, such as generalized estimating equations (GEE) and random-effects models (REM), provide robust solutions for handling clustering.[21] A simpler alternative is to use cluster sampling theory, where the prevalence of UDC is calculated as the average percentage across individuals.[22]

Despite their availability, these methods have not been compared in the context of dental data. Therefore, this study compared different methods for estimating the prevalence of untreated dental caries to identify the most accurate and practical approach for dental public health applications.

Methods

Overview of Estimating the Prevalence of Untreated Dental Caries

This study evaluated four methods for estimating the prevalence of UDC: (1) overall prevalence, (2) average individual prevalence, (3) GEE, and (4) REM. These methods were selected to account for different levels of complexity in prevalence estimation, ranging from simple aggregation to advanced statistical modeling that incorporates clustering effects. Supplementary 1 provides technical details for each method.

To illustrate these methods, we used a hypothetical sample of five individuals with varying numbers of teeth and untreated decayed teeth (Table 1). This example demonstrates the calculation of prevalence using each method.

Overall Prevalence Estimator

This method calculates prevalence by dividing the total number of decayed teeth by the total number of teeth across all individuals (15% in Table 1). Confidence intervals are calculated using normal approximation (see Supplementary 1). While simple, this method does not account for individual-level variability or clustering effects.

Average Individual Prevalence Estimator

This method calculates the prevalence for each individual and then averages these values (e.g., the average of the fourth column in Table 1 is 21.7%). Confidence intervals are estimated using the bootstrap method, which avoids the complexity of analytically deriving standard errors. This approach accounts for both individual-level variability and clustering effects.

GEE Estimator

Estimating prevalence using the GEE method requires transforming count data into a binary dataset formatted in long format. In this format, each record indicates the status of each tooth: 0 represents another tooth, and 1 represents a UDC tooth. A personal identifier is included to group teeth by subject. The database may also contain additional variables linked to each tooth, such as tooth number, and to each subject, including age and gender (Table 2). Various analyses, including logistic regression models, can be applied to the data, accounting for the clustering of teeth within individuals using the personal identifier.

Once the data are converted to binary format, logistic regression is used to estimate prevalence. In this context, participant IDs are specified in the statistical software using the GEE approach for the logistic regression model, which treats individuals as clusters and accounts for the clustering effect in the prevalence estimation (see Supplementary 1 for further details). Ultimately, this methodology ensures that the clustering effect is appropriately incorporated into the estimation process, enhancing the accuracy of the prevalence calculations.

REM Estimator

In the long dataset of UDC, we may evaluate the

Table 1. Hypothetical example of sampling five subjects from a population with an unknown prevalence of dental caries

Subject's ID	Total number of teeth	Untreated decayed teeth	Percentage of untreated decayed teeth in each subject
1	20	2	10%
2	26	1	4%
3	7	3	43%
4	10	4	40%
5	17	2	12%
Total	80	12	Average individual prevalence = 21.7% Overall prevalence = 15%

prevalence of UDC at both the individual and tooth levels. In this case, REM models, also known as multilevel models, estimate prevalence at two levels. For tooth-level prevalence, a known fixed value is assumed; whereas at the individual level, a distribution with a mean of zero and an unknown variance is assumed. Ultimately, complex statistical techniques are used to estimate the variance of prevalence at the individual and tooth levels, along with their confidence intervals (for more details, see Supplementary 1).

Table 2. Long format of hypothetical example for GEE and REM methods

Subject's ID	Number of teeth	Condition (Tooth with untreated dental caries = 1 and other tooth = 0)
1	1	1
1	2	1
1	3	0
.	.	.
.	.	.
.	.	.
.	.	.
1	20	1
2	1	0
2	2	0
2	3	0
2	4	0
.	.	.
.	.	.
.	.	.
.	.	.
2	26	0
3	1	1
3	2	1
3	3	1
3	4	0
3	5	0
3	6	0
3	7	0
.	.	.
.	.	.
.	.	.
.	.	.

Simulation Study

To compare the performance of the four prevalence estimation methods, we conducted a simulation study. Hypothetical populations were generated with specified UDC prevalence levels and varying degrees of correlation in untreated caries risk among individuals, measured by the intraclass correlation coefficient (ICC). Figure 1 presents the simulation process.

Hypothetical Populations and Generation

The first step of our simulation study was to create a reference population. To create the reference population, 5000 random numbers were selected from a uniform discrete distribution ranging from 14 to 28, representing the number of teeth for individuals in the population. We considered two scenarios to generate hypothetical populations with varying prevalences and ICC values:

1. Untreated dental caries independent of missing teeth

In the first part of the scenarios, UDC prevalence was fixed at 5%, 10%, or 20%, with ICC values of 0.05, 0.1, and 0.2. These combinations yielded 9 populations, assuming equal prevalence across all individuals (Table 3). This scenario reflects situations in which the number of missing teeth does not affect UDC prevalence.

2. Untreated Dental Caries Dependent on Missing Teeth

In the second part of the scenarios, UDC prevalence increased with the number of missing teeth. For individuals with 28 teeth, the prevalence was set at 5%, increasing by 1% for each missing tooth (e.g., for M missing teeth, prevalence = (5% + 1%) × M). This dependency was modeled across ICC values of 0.05, 0.1, and 0.2, resulting in three additional populations (Table 3).

Hypothetical Populations Generation

To generate hypothetical populations, the number of

Table 3. Prevalence of UDC and ICC in hypothetical population scenarios

	Hypothetical populations	Prevalence of untreated dental caries in individuals	Interclass correlation (similarity of an individual's teeth)
Untreated dental caries Independent of missing teeth	Population 1	5%	0.05
	Population 2	5%	0.1
	Population 3	5%	0.2
	Population 4	10%	0.05
	Population 5	10%	0.1
	Population 6	10%	0.2
	Population 7	20%	0.05
	Population 8	20%	0.1
	Population 9	20%	0.2
Untreated dental caries dependent on missing teeth	Population 10	(5% + 1%) × the number of missing teeth for each individual	0.05
	Population 11	(5% + 1%) × the number of missing teeth for each individual	0.1
	Population 12	(5% + 1%) × the number of missing teeth for each individual	0.2

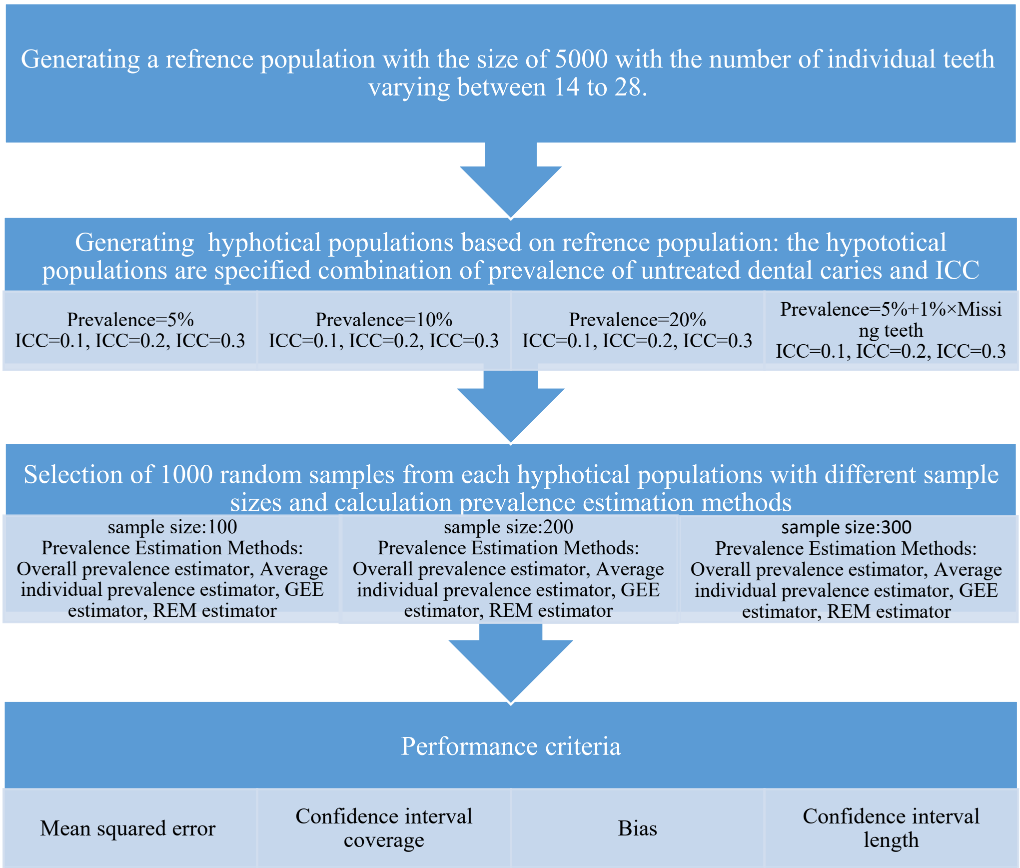

Figure 1. Flowchart of simulation study steps to compare four different methods of untreated dental caries estimation

untreated decayed teeth per person in the reference population was generated from the beta-binomial distribution. This distribution was particularly used to generate clustered data when the interclass correlation was specified at specific values.[20,21] To generate data from this distribution, three parameters had to be determined: the number of teeth for each individual in the population, the prevalence of UDC, and the interclass correlation. Using the beta-binomial distribution ensured that the prevalence of UDC in the generated population matched the specified values. However, due to the inherent randomness in the number generation process, the actual prevalence in the generated population was not exactly equal to the set value; the difference was very small, reflecting the natural variability associated with random sampling. After generating the population, we calculated the actual prevalence rate using the resulting data. This calculated prevalence was then used in our subsequent analyses and computations to ensure that our results accurately reflected the characteristics of the generated population.

Sampling and Estimation

We sampled sizes of 100, 200, and 300 from hypothetical populations without replacement, repeating this process 1,000 times for each sample. To estimate prevalence using the GEE method, the “geepack” package was used; for the REM method, the “lme4” package was used to fit the sampled data. For the cluster estimator, the package “boot” was used to estimate the prevalence confidence interval. Also, the simple estimator was used to estimate the prevalence of UDC as described in the previous section, with the syntax in the R software. All simulations and calculations were conducted using R statistical software (v4.4.2; R Core Team 2024).

Comparison Criteria Calculation

The prevalence estimation method focused on two key metrics: mean squared error (MSE) and confidence interval coverage, selected to evaluate the accuracy of the estimator and the reliability of the confidence intervals, respectively. Additionally, bias and the length of the confidence interval were analyzed, with detailed results provided in Supplementary 2. The metrics were defined and calculated as follows:

1. **Mean squared error (MSE):** The MSE of each estimator was assessed as the average squared difference between the estimated and true prevalences across simulations.
2. **Confidence interval coverage:** The coverage of the confidence intervals was evaluated to determine the proportion of intervals that contained the true prevalence value.
3. **Confidence interval length:** The confidence interval length was used to assess the precision of the estimators. Shorter intervals indicate greater precision.
4. **Bias:** The bias of each estimator was assessed by calculating the average difference between the estimated and true prevalences across simulations.

Supplementary 2 presents all R code related to the simulations and calculations, as well as the results.

Empirical Study

To demonstrate the practical utility of our simulation-based evaluation, we applied the results of our simulation study to data from the third phase of the Kerman Coronary Artery Disease Risk Factors Study (KERCADRS), a population-based cohort conducted in Kerman Province, Iran. The study adhered to the Declaration of Helsinki, and its protocol was approved by the Ethics Committee of Kerman University of Medical Sciences, Iran (Ethical code: IR.KMU.REC.1399.609). The study was conducted among residents of Kerman city aged 15–75 years, and complete information on the sampling method and participant recruitment is available in the references.[23,24] For this analysis, we included participants with completed oral health assessments in Phase III and with complete data on the a priori-specified covariates. We excluded the participants without any natural teeth.

The primary outcome was untreated dental caries at the person level, defined as the total number of decayed teeth and the total number of teeth identified by trained examiners during clinical examination under standardized conditions. Consistent with WHO oral health survey criteria, a tooth was classified as "untreated carious" if active caries was present without evidence of definitive restorative or endodontic care. We derived a binary indicator at the individual level (any untreated decay: yes/no). Auxiliary covariates considered for adjustment or modeling included age and sex.

Results

In this section, we first present a hypothetical example of sampling five subjects from a population with an unknown prevalence of dental caries, and then present the results of a simulation study.

Hypothetical Example of Data Prevalence Estimation

The estimated prevalence of untreated dental caries among a sample of five hypothetical individuals varied across different estimation methods. The overall prevalence was 15%, while the GEE method produced a higher estimate of 18.5%. The REM estimate was slightly lower at 15.6%. The average individual prevalence was the highest at 21.6%. Notably, the confidence intervals for these estimates also differed, reflecting the variability in the prevalence estimates across the methods (Table 4).

Table 4. Estimated prevalence of untreated dental caries in a sample of 5 hypothetical individuals described in the methods section using various estimation methods

Estimation method	Estimated prevalence (%)	Confidence interval (95%)
Overall	15	(7.1, 22.8)
GEE estimate	18.5	(8.7, 34.9)
REM estimate	15.6	(7.3, 30.5)
Average individual	21.6	(7.8, 35.7)

Simulation Study Results

Untreated dental caries independent of missing teeth

When the number of UDC was held constant and independent of the number of missing teeth within the population, the REM estimator exhibited the highest MSE, making it the least accurate among the estimators assessed. Conversely, all alternative estimators yielded identical MSE values (Table 5), indicating consistent accuracy across the approaches. However, the confidence interval coverage for overall prevalence and the REM estimator consistently fell below 85% in all tested scenarios, demonstrating a significant failure of these methods to produce reliable confidence intervals. Notably, the REM estimator exhibited the lowest coverage, emphasizing its limitations. The average individual prevalence and GEE estimator achieved nearly 95% coverage, particularly at a sample size of 300, where they attained at least 95% coverage, thereby demonstrating superior performance and more reliable confidence intervals (Table 6). Specifically, in the most extreme scenario—characterized by the lowest prevalence of UDC (5%), the highest ICC (0.2), and the smallest sample size (100)—both the GEE and average individual prevalence estimators demonstrated the most robust confidence interval coverage, indicating their reliability in estimating true values under challenging conditions.

Regarding bias, the GEE estimator, overall prevalence, and average individual prevalence exhibited equivalent average bias values (Supplementary file, 2, Table S1). Furthermore, both the GEE and average individual prevalence estimators produced identical confidence interval lengths (Supplementary file, 2, Table S2).

Treatment of Untreated Dental Caries Relative to Missing Teeth

When the number of UDC depended on the number of missing teeth, the REM method again yielded the highest MSE, confirming its status as the least accurate estimator of UDC prevalence. In contrast, the overall prevalence, GEE, and average individual prevalence estimators yielded very similar MSE values (Table 5). However, the overall prevalence method achieved the lowest MSE. However, it is noteworthy that the confidence interval coverage for overall prevalence remained below 85%. In comparison, the GEE and average individual prevalence methods showed higher coverage, but remained below 95% (Table 6). In the most extreme scenario, characterized by the highest ICC (0.2) and the smallest sample size (100), the GEE method exhibited the best confidence interval coverage.

In terms of bias, the overall prevalence, GEE estimator, and average individual prevalence showed the lowest bias when the number of UDC depended on the number of missing teeth in the population (Table 1, Supplementary file, 1). Furthermore, both the GEE and average individual prevalence estimators maintained identical confidence interval lengths (Table 2, Supplementary file, 1).

Empirical Study Results

To estimate UDC in phase three of the KERCADRS

Table 5. Mean square error (MSE) of individual prevalence, GEE estimator, overall prevalence, and REM estimator for untreated dental caries across varying sample sizes and ICC values

Prevalence	ICC	Sample size	Average individual prevalenceestimator	GEE estimator	Overall prevalenceestimator	REM estimator
0.05	0.05	100	0.00005	0.00005	0.00005	0.00027
		200	0.00002	0.00002	0.00002	0.00026
		300	0.00001	0.00001	0.00001	0.00026
	0.1	100	0.00007	0.00007	0.00007	0.00073
		200	0.00004	0.00003	0.00003	0.00073
		300	0.00002	0.00002	0.00002	0.00073
	0.2	100	0.00011	0.00011	0.00011	0.00174
		200	0.00006	0.00006	0.00006	0.00174
		300	0.00003	0.00003	0.00004	0.00176
0.1	0.05	100	0.00009	0.00008	0.00008	0.00039
		200	0.00004	0.00004	0.00004	0.00036
		300	0.00003	0.00003	0.00003	0.00034
	0.1	100	0.00012	0.00012	0.00013	0.00112
		200	0.00006	0.00006	0.00006	0.00107
		300	0.00004	0.00004	0.00004	0.00105
	0.2	100	0.00020	0.00020	0.00021	0.00347
		200	0.00009	0.00009	0.00009	0.00343
		300	0.00006	0.00006	0.00006	0.00342
0.2	0.05	100	0.00014	0.00014	0.00014	0.00039
		200	0.00008	0.00008	0.00008	0.00032
		300	0.00005	0.00005	0.00005	0.00026
	0.1	100	0.00023	0.00023	0.00023	0.00111
		200	0.00010	0.00010	0.00010	0.00103
		300	0.00008	0.00007	0.00007	0.00099
	0.2	100	0.00036	0.00036	0.00036	0.00390
		200	0.00017	0.00017	0.00017	0.00373
		300	0.00012	0.00012	0.00013	0.00372
5% + 1% × number of missing teeth for each individual	0.05	100	0.00028	0.00022	0.00020	0.00058
		200	0.00019	0.00013	0.00011	0.00048
		300	0.00014	0.00009	0.00007	0.00042
	0.1	100	0.00036	0.00032	0.00029	0.00136
		200	0.00023	0.00019	0.00014	0.00122
		300	0.00017	0.00013	0.00009	0.00112
	0.2	100	0.00052	0.00050	0.00050	0.00410
		200	0.00028	0.00026	0.00025	0.00384
		300	0.00020	0.00019	0.00016	0.00383

study, data from 6,016 individuals were included in the final analysis. The data showed a significant correlation between the number of UDC and the number of missing teeth ($r = 0.15$, $P < 0.001$). Accordingly, the prevalence of UDC estimated using the GEE method was 26.3% (95% CI: 25.9%, 26.8%). There was no significant difference in UDC prevalence between the two sexes ($P = 0.204$). With each 1-year increase in participants' age, the odds of UDC increased by 1.01 ($P < 0.001$).

Discussion

Accurate estimation of UDC prevalence is essential for assessing public health trends and guiding intervention strategies. Our study is the first to evaluate different methods for estimating prevalence under varying conditions, with a particular focus on the relationship between UDC and the number of missing teeth. The results of our study indicate that when the number of UDC was independent of the number of missing teeth, both the GEE method and individual average prevalence provided the most reliable estimates. However, when the number of UDC depended on the number of missing teeth, none of the methods performed optimally, though the GEE method performed relatively better. These findings highlight the limitations of prevalence estimation when there is a dependency between the number of UDC

Table 6. Confidence interval coverage for individual prevalence, GEE estimator, overall prevalence, and REM estimator of untreated dental caries across different sample sizes and ICC values

Prevalence	ICC	Sample size	Average individual prevalenceestimator	GEE estimator	Overall prevalenceestimator	REM estimator
0.05	0.05	100	0.94	0.94	0.83	0.42
		200	0.96	0.96	0.84	0.09
		300	0.95	0.95	0.85	0.02
	0.1	100	0.94	0.94	0.73	0.06
		200	0.95	0.95	0.75	0.00
		300	0.95	0.95	0.74	0.00
	0.2	100	0.93	0.94	0.59	0.00
		200	0.93	0.93	0.60	0.00
		300	0.95	0.95	0.63	0.00
0.1	0.05	100	0.94	0.95	0.84	0.38
		200	0.94	0.95	0.83	0.12
		300	0.95	0.95	0.83	0.03
	0.1	100	0.94	0.95	0.73	0.19
		200	0.95	0.96	0.74	0.01
		300	0.95	0.95	0.76	0.00
	0.2	100	0.95	0.95	0.63	0.01
		200	0.96	0.96	0.67	0.00
		300	0.96	0.96	0.65	0.00
0.2	0.05	100	0.94	0.95	0.84	0.80
		200	0.94	0.95	0.83	0.59
		300	0.96	0.96	0.85	0.51
	0.1	100	0.94	0.94	0.74	0.54
		200	0.95	0.95	0.74	0.22
		300	0.95	0.95	0.72	0.08
	0.2	100	0.95	0.95	0.63	0.19
		200	0.97	0.96	0.65	0.02
		300	0.96	0.95	0.60	0.00
5% + 1% × number of missing teeth for each individual	0.05	100	0.91	0.94	0.82	0.74
		200	0.85	0.92	0.80	0.57
		300	0.81	0.92	0.81	0.43
	0.1	100	0.91	0.93	0.73	0.56
		200	0.87	0.91	0.75	0.28
		300	0.84	0.91	0.74	0.13
	0.2	100	0.93	0.94	0.60	0.33
		200	0.92	0.93	0.58	0.08
		300	0.91	0.93	0.61	0.01

and the number of missing teeth.

In the present study, the performance of the REM estimator differed from that of GEE, overall prevalence, and individual average prevalence in terms of MSE and bias. Consistent with our results, Zhang et al. also noted differences in regression coefficient estimates between the REM and GEE under various conditions.[25] We also applied the anti-logit transformation to the regression coefficient to estimate prevalence, which can amplify minor differences in regression coefficients into larger differences in prevalence. Therefore, the use of REM for estimating UDC prevalence is not recommended.

We showed that although the overall prevalence estimator had bias and MSE similar to those of the GEE method, its confidence intervals had low coverage and were short. This phenomenon arises from the neglect of the ICC in variance calculations, leading to an underestimation of the true variance and, consequently, to erroneous statistical inferences.[26] Consequently, relying on confidence intervals of the overall prevalence estimator is not advisable, even with weak dependency in UDC among study participants.

Our results indicated that, in the absence of correlation between the number of decayed and missing teeth, both the GEE and individual prevalence methods produced valid and comparable estimates with associated confidence intervals. However, in the presence of a strong correlation, these methods failed to yield valid estimates or confidence intervals. The alternative approach could be considering the number of missing teeth as an independent variable within the model. If the coefficient for this variable is found to be significant, it is essential to report the baseline prevalence and the coefficient of the number of missing teeth, which accurately reflects the prevalence of UDC in individuals without missing teeth. However, the performance of this approach should be evaluated in future studies.

A strength of our study was its comprehensive comparison of prevalence estimation methods, as it considered multiple parameters to assess their performance. However, its main limitation was that we employed the beta-binomial model to generate dependent data, noting that different results may arise when using other models for dependent binary data. Thus, future research should explore various prevalence estimation methods with different dependency structures. Also, future research should explore the level of dependence between decayed and missing teeth at which including the number of missing teeth in the model becomes unnecessary. Additionally, further studies should explore these findings in the context of higher prevalence rates to enhance generalizability.

Our empirical findings align with global evidence: untreated dental caries (UDC) tends to progress and can lead to tooth loss if care is delayed, making UDC a useful early warning sign of service need.[27] The lack of a clear difference between men and women suggests that shared, changeable factors—like high sugar intake, irregular use of dental services, and barriers to access—are more important than biology, so broad, population-wide prevention is likely to work best.[27,28] The increase in risk with age is consistent with lifelong build-up of exposures and gaps in prevention, which means we need steady, age-appropriate prevention across the life course.

Conclusion

Our findings highlight the importance of choosing appropriate statistical methods for estimating dental caries prevalence. The GEE method and individual average prevalence provided the most accurate estimates when there was no association between the number of untreated dental caries and missing teeth. However, when dependencies exist between these variables, it is advisable to include the number of missing teeth in the GEE model and to report the baseline prevalence along with the coefficient of missing teeth.

Acknowledgements

This research was part of the PhD thesis of Amir Hossein Nekouei We would like to express our sincere gratitude to all those who contributed to this work, including colleagues and mentors.

Authors' Contribution

Conceptualization: Aliakbar Haghdoost, Mohammad Reza Baneshi
Data curation: Amir Hossein Nekouei
Formal analysis: Amir Hossein Nekouei
Funding acquisition: Aliakbar Haghdoost.
Investigation: All authors
Methodology: Aliakbar Haghdoost, Mohammad Reza Baneshi
Project administration: Aliakbar Haghdoost, Mohammad Reza Baneshi
Software: Amir Hossein Nekouei
Supervision: Aliakbar Haghdoost, Mohammad Reza Baneshi
Validation:All authors
Writing–original draft: All authors
Writing–review & editing: All authors

Competing Interests

The authors declare that they have no competing interests.

Consent to Participate

Not applicable.

Consent to Publish

Not applicable.

Ethical Approval

Not applicable.

Funding

The study was partially funded by Kerman Univercity of Medical Sciences University of Medical Sciences(Fund No.: 402000291).

Supplementary File

Supplementary file 1. R codes
Supplementary file 2. Bias and confidence interval length comparision

References

1. Davidson T, Bergström EK, Husberg M, Moberg Sköld U. Long-Term Cost-Effectiveness through the Dental-Health FRAMM Guideline for Caries Prevention. Int J Environ Res Public Health 2022;19(4):1954. doi:10.3390/ijerph19041954
2. Janusz CB, Doan TT, Gebremariam A, Rose A, Keels MA, Quinonez RB, et al. A Cost-Effectiveness Analysis of Population-Level Dental Caries Prevention Strategies in US Children. Acad Pediatr 2024;24(5):765–75. doi:10.1016/j.acap.2024.02.006
3. Johhnson B, Serban N, Griffin PM, Tomar SL. Projecting the economic impact of silver diamine fluoride on caries treatment expenditures and outcomes in young U.S. children. J Public Health Dent 2019;79(3):215–21. doi:10.1111/jphd.12312
4. Weintraub JA, Stearns SC, Rozier RG, Huang CC. Treatment outcomes and costs of dental sealants among children enrolled in Medicaid. Am J Public Health 2001;91(11):1877–81. doi:10.2105/ajph.91.11.1877
5. Ono S, Sasabuchi Y, Ishimaru M, Ono Y, Matsui H, Yasunaga H. Short-term effects of reduced cost sharing on childhood dental care utilization and dental caries prevention in Japan. Community Dent Oral Epidemiol 2023;51(2):228–35. doi:10.1111/cdoe.12730
6. Elamin A, Garemo M, Mulder A. Determinants of dental caries in children in the Middle East and North Africa region: a systematic review based on literature published from 2000 to 2019. BMC Oral Health 2021;21(1):237. doi:10.1186/s12903-021-01482-7
7. Organization WH. Oral health surveys: basic methods: World Health Organization; 2013. Accessed May 5, 2026. Available from: https://iris.who.int/bitstream/handle/10665/97035/9789241548649_eng.pdf
8. Angelillo IF, Anfosso R, Nobile CG, Pavia M. Prevalence of dental caries in schoolchildren in Italy. Eur J Epidemiol 1998;14(4):351–7. doi:10.1023/a:1007471707836
9. Teshome A, Muche A, Girma B. Prevalence of Dental

Caries and Associated Factors in East Africa, 2000-2020: Systematic Review and Meta-Analysis. Front Public Health 2021;9:645091. doi:10.3389/fpubh.2021.645091

10. Eid SA, Khattab NMA, Elheeny AAH. Untreated dental caries prevalence and impact on the quality of life among 11 to14-year-old Egyptian schoolchildren: a cross-sectional study. BMC Oral Health 2020;20(1):83. doi:10.1186/s12903-020-01077-8
11. Murthy AK, Pramila M, Ranganath S. Prevalence of clinical consequences of untreated dental caries and its relation to dental fear among 12-15-year-old schoolchildren in Bangalore city, India. Eur Arch Paediatr Dent 2014;15(1):45–9. doi:10.1007/s40368-013-0064-1
12. Nath S, Sethi S, Bastos JL, Constante HM, Mejia G, Haag D, et al. The Global Prevalence and Severity of Dental Caries among Racially Minoritized Children: A Systematic Review and Meta-Analysis. Caries Res 2023;57(4):485–508. doi:10.1159/000533565
13. Ntani G, Inskip H, Osmond C, Coggon D. Consequences of ignoring clustering in linear regression. BMC Med Res Methodol 2021;21(1):139. doi:10.1186/s12874-021-01333-7
14. Galbraith S, Daniel JA, Vissel B. A study of clustered data and approaches to its analysis. J Neurosci 2010;30(32):10601–8. doi:10.1523/jneurosci.0362-10.2010
15. Siraneh Y, Woldie M, Birhanu Z. Ignoring Clustering and Nesting in Cluster Randomized Trials Renders Conclusions Unverifiable [Response to Letter]. Risk Manag Healthc Policy 2022;15:2011–4. doi:10.2147/rmhp.S392171
16. Wilson J, Lorenz K. Generalized Estimating Equations Logistic Regression. Modeling Binary Correlated Responses using SAS, SPSS and R. Springer International Publishing 2015. p. 103–30. doi:10.1007/978-3-319-23805-0_6
17. Wu C, Thompson M. Stratified Sampling and Cluster Sampling. Sampling Theory and Practice. Springer International Publishing 2020. p. 33–56. doi:10.1007/978-3-030-44246-0_3
18. Najafipour H, Mirzazadeh A, Haghdoost A, Shadkam M, Afshari M, Moazenzadeh M, et al. Coronary Artery Disease Risk Factors in an Urban and Peri-urban Setting, Kerman, Southeastern Iran (KERCADR Study): Methodology and Preliminary Report. Iran J Public Health 2012;41(9):86–92.
19. Physiology Research Center IoN, Kerman University of Medical Sciences, Kerman, Iran. Kerman Coronary Artery Diseases Riskfactor Study-KERCADRS. Physiology Research Center IoN. Accessed March 2, 2026. Available from: https://kprc.kmu.ac.ir/en/page/21297/Introduction
20. Zhang H, Xia Y, Chen R, Gunzler D, Tang W, Tu X. Modeling longitudinal binomial responses: implications from two dueling paradigms. Journal of Applied Statistics 2011;38(11):2373–90. doi:10.1080/02664763.2010.550038
21. Julian M. The Consequences of Ignoring Multilevel Data Structures in Nonhierarchical Covariance Modeling. Structural Equation Modeling-a Multidisciplinary Journal - STRUCT EQU MODELING 2001;8:325–52. doi:10.1207/S15328007SEM0803_1
22. Watt RG, Daly B, Allison P, Macpherson LMD, Venturelli R, Listl S, et al. Ending the neglect of global oral health: time for radical action. Lancet 2019;394(10194):261–72. doi:10.1016/s0140-6736(19)31133-x
23. Peres MA, Macpherson LMD, Weyant RJ, Daly B, Venturelli R, Mathur MR, et al. Oral diseases: a global public health challenge. Lancet 2019;394(10194):249–60. doi:10.1016/s0140-6736(19)31146-8
24. Weintraub JA, Ramos-Gomez F, Jue B, Shain S, Hoover CI, Featherstone JD, et al. Fluoride varnish efficacy in preventing early childhood caries. J Dent Res 2006;85(2):172–6. doi:10.1177/154405910608500211
25. Griffin SO, Lin M, Scherrer CR, Naavaal S, Hopkins DP, Jones AA, et al. Effectiveness of School Fluoride Delivery Programs: A Community Guide Systematic Review. Am J Prev Med 2025;69(1):107633. doi:10.1016/j.amepre.2025.04.003
26. Chamut S, Alhassan M, Hameedaldeen A, Kaplish S, Yang AH, Wade CG, et al. Every bite counts to achieve oral health: a scoping review on diet and oral health preventive practices. Int J Equity Health 2024;23(1):261. doi:10.1186/s12939-024-02279-0
27. Vernon LT, Teng KA, Kaelber DC, Heintschel GP, Nelson S. Time to integrate oral health screening into medicine? A survey of primary care providers of older adults and an evidence-based rationale for integration. Gerodontology 2022;39(3):231–40. doi:10.1111/ger.12561
28. Shrivastava R, Couturier Y, Girard F, Papineau L, Emami E. Two-eyed seeing of the integration of oral health in primary health care in Indigenous populations: a scoping review. Int J Equity Health 2020;19(1):107. doi:10.1186/s12939-020-01195-3