ST open

RESEARCH ARTICLE

Estimating the Decayed, Missing, and Filled permanent Teeth (DMFT) index of the University of Split student population using the *Monte Carlo* method

Dora Dodig, Darko Kero

University of Split School of Medicine, Graduate Study of Dental Medicine

Aim: To generate data for the Decayed, Missing, and Filled Permanent Teeth (DMFT) index of the entire student population of the University of Split from three samples of different sizes using the Monte Carlo method.

Methods: We collected data during clinical exercises in the courses 'Restorative Dental Medicine 2' and 'Endodontics 2.' All participants (n = 200) were students at the University of Split. We derived three samples from the collected data on the DMFT index – small (n = 50), medium-sized (n = 100), and the entire sample (n = 200). Afterwards, we ran Monte Carlo simulations (MCS) to derive DMFT index data for three student populations of 20 000 individuals each from those samples. The probabilities of individual outcomes for the DMFT index and its components, as well as their correlations, were entered as assumptions for the MCS.

Results: The estimated mean DMFT index of the student population was 8.96 (standard deviation (SD) = 0.69, 99% CI = 8.91–9.00) for MCS 50, 9.12 (SD = 0.47, 99% CI = 9.08–9.21) for MCS 100, and 8.82 (SD = 0.36, 99% CI = 8.77–8.87) for MCS 200. Regarding the components of the DMFT index, the number of repaired teeth in MCS 50, MCS 100, and MCS 200 was most strongly correlated with the DMFT index, with R values of 0.84, 0.82, and 0.77, respectively. The estimated mean DMFT indices by MCS differed from the mean DMFT indices of the corresponding samples and each other by less than 1 point.

Conclusions: The Monte Carlo method may be useful in estimating the population means of clinical indices in dental medicine, including the DMFT index. According to the assumptions made for MCS in this study, the size of the original samples did not significantly affect the estimates of the parameters of the DMFT index.

Keywords: DMFT; Monte Carlo method; oral health; population; sampling; simulation

Correspondence to:
Darko Kero
University of Split School of Medicine,
Šoltanska 2, 21000 Split, Croatia
dkero@mefst.hr

Cite as:
Dodig D, Kero D. Estimating the Decayed, Missing, and Filled permanent Teeth (DMFT) index of the University of Split student population using the Monte Carlo method. ST-OPEN. 2024;5:e2024.2308.7.

DOI:
https://doi.org/10.48188/so.5.2

Introduction

One of the most important prerequisites for the validity of statistical analyses in scientific research is the representativeness of the sample in relation to the population under investigation (*1*). If the sample is representative, the parameters of the measured variable(s) within the study sample can be more accurately generalized to the total population from which the sample was taken.

When creating a representative sample, a researcher has to determine the appropriate sample size (*2*). Researchers may be guided by the simple logic that the larger the sample, the more representative it is of the population from which it is drawn when an appropriate sampling method is used. However, taking and processing a large sample (e.g., in experimental laboratory research and clinical research) is often not feasible due to financial and organizational restrictions. Therefore, statistical procedures were designed as part of the power analysis with which the required sample size can be determined *a priori* (*3*). Another way to ensure the sample's representativeness is to introduce randomness in a population sample, whereby each individual from a given population has an equal chance of being included in the study (*1*, *4*). It is also easier to ensure sample representativeness when certain variables are expressed as indices, which is convenient for epidemiologic and clinical research (*5*). For example, in dental medicine, various clinical indices are used to measure oral health (which is a complex variable in itself). The best-known among them is the Decayed, Missing, and Filled permanent Teeth (DMFT) index as the sum of a person's decayed, extracted and repaired permanent teeth with a theoretically defined range from 0 to 64 (*6*-*9*).

The shortcoming of previously described methods used to achieve sample representativeness is that, if a particular population parameter is not known, it can never be said with certainty that sample representativeness has been achieved through their application (*10*, *11*). For example, the distribution of values for a variable in the sample, regardless of whether it is a uniform distribution, normal distribution, or some type of asymmetric/skewed distribution, does not have to correspond to the actual distribution in a population (*12*). Another challenge emerges with oral diseases; for example, in the case of caries, researchers face sampling problems due to its generally high prevalence, which also varies between age groups. It is therefore logical to assume that, to estimate the prevalence of dental caries and its consequences in the population, researchers have to work only with large samples or need to repeat the same studies several times on new samples taken from a large population. The purpose of both approaches is to increase the sample size to allow a better statistical evaluation of the population parameters of the disease.

An alternative approach for achieving sample representativeness is based on a stochastic algorithmic procedure according to the Monte Carlo method (*13*). In this approach, repeated sampling can be simulated *in silico* by using random or quasi-random numbers and a large number of calculations, taking into account that the parameters of the simulated samples differ from each other, reflecting the diversity of the population under study; that, according to the Central Limit Theorem, the larger the number of simulated samples, the distribution of the mean values of the measured variable in the simulated samples be-

comes more normal; and that the central place in this normal distribution belongs to the mean value of the measured variable that is close to the true population mean (*12*).

This study aimed to estimate the population mean of the DMFT index of the University of Split student population by generating DMFT index data for the entire University of Split student population from three samples of different sizes using the Monte Carlo method.

Methods

Study participants and data collection

We collected data in the academic years 2021/22 and 2022/23 during the clinical exercises in the courses 'Restorative Dentistry 2' and 'Endodontics 2.' We recruited study participants using convenience sampling from students at the University of Split who came for an examination or treatment to the Polyclinic of the Study of Dental Medicine 'Dental academicus' in the premises of the student dormitory 'Dr. Franjo Tuđman' in Split. The study participants signed an informed consent form for treatment and the processing of personal data for research purposes. We recorded their personal, family, and dental history, performed a clinical examination of the oral cavity, and recorded the oral status using the DMFT index. The examinations were performed by fourth- and fifth-year dental students at the University of Split School of Medicine, under the supervision of the course instructors. The Ethics Committee of the University of Split School of Medicine approved the study (Class 003-08//23-03/0015; No. 2181-198-03-04-23-0082). The raw data and data simulated using the Monte Carlo method were stored in a spreadsheet in Microsoft Office Excel 2016 (Microsoft Corporation, Redmond, WA, USA).

Statistical analysis

We analyzed demographic parameters such as age and gender using descriptive statistics, expressing them as numerical values, percentages and/or mean values with standard deviations (SDs) and 99% confidence intervals (CIs). We also presented the scores for the DMFT index and the sub-indices (decayed teeth (D), missing teeth (M), and filled teeth (F) numerically (discrete data) and as mean scores with SDs and 99% CIs. We presented the data distribution of all variables (except gender) through histograms.

We then derived three samples from the data collected on the DMFT index – a small sample of the first 50 participants, a medium sample of the next 100 participants, and the entire sample of 200 participants. We defined two main criteria for simulating the samples with Monte Carlo simulation (MCS) and generating data for the entire student population: the probabilities of the individual outcomes for DMFT and DMFT components (D, M, and F), and the correlation between the DMFT index and DMFT components. The probabilities of the individual outcomes are represented by the probability density function (PDF) and the cumulative distribution function (CDF). Since the scores of the DMFT index and its components are integers, only the latter representation was introduced as a criterion for the MCS. Although the theoretical range of the DMFT index is from 0 to 64, the outcome prob-

abilities are shown for the range from 0 to 41 after the predicted number of iterations for MCS from all three samples have been performed. To include the correlation of the DMFT index and its components as an additional criterion for MCS, we created correlation matrices for the mentioned variables for each of the three samples (Pearson's correlation). After calculating the matrix determinants, we decomposed these correlation matrices using the Cholesky decomposition. Since the DMFT index can only have positive integer values in the range 0-64, we made two subsequent corrections to the decimal and negative DMFT scores generated in MCS, i.e., rounding decimal numbers to integer values and replacing negative values with zero. The decimal and negative DMFT scores appeared due to the multiplication of matrices with raw data with decomposed correlation matrices, which was necessary to introduce the correlation between the DMFT index and the sub-indices D, M and F as one of the assumptions of MCS. We recorded the values and parameters of the variables in the simulated samples and the generated populations (means, SDs, 99% CIs, ranges, and correlation coefficients) in separate spreadsheets.

We explored the correlation between the DMFT index and its components in the simulated student populations by simple linear regression. with the DMFT index as the main outcome and the D, M, and F sub-indices as predictors. We then analyzed the differences in the mean values of the DMFT index of the simulated samples generated in MCS by multiple linear regression with the DMFT index as the outcome and the predictors coded as categorical 'dummy' variables corresponding to the small sample (MCS 50), the medium sample (MCS 100) and the whole sample (MCS 200). The dummy variables were coded as described previously (*14*, *15*).

We performed all statistical analyses entirely in Microsoft Excel 2016 (Microsoft Corporation, Redmond, WA, USA). The significance level was set at $\alpha = 0.01$ ($p < 0.01$).

Results

Demographic parameters of the participants and parameters of the DMFT index of the samples

We collected data on the DMFT index from participants (n = 200) with the mean age of 22.83 years (SD = 2.80, 99% CI = 22.57–23.08). The gender distribution in the entire sample was relatively balanced with 46.50% (n/N = 93/200) men and 53.50% (n/N = 107/200) women. The mean age of the subjects in the small sample (n = 50) was 26.36 years (SD = 2.24, 99% CI = 25.93–26.79) with a gender distribution of 58% (n/N = 29/50) men and 42% (n/N = 21/50) women. The mean age of the subjects in the median sample (n = 100) was 22.29 years (SD = 0.84, 99% CI = 22.18–22.41) with an even gender distribution of 50% men (n/N = 50/100) and women (n/N = 50/100).

The parameters of the DMFT index and the components of the DMFT index based on the data from the sample of the student population of the University of Split are presented in **Table 1.**

Table 1. Descriptive statistics of DMFT index and DMFT index components from samples of the student population of the University of Split*

	Small sample (n = 50)		Medium sample (n = 100)		Entire sample (n = 200)	
Parameters	Mean (SD)	99% CI	Mean (SD)	99% CI	Mean (SD)	99% CI
D	2.42 (2.31)	1.98–2.86	2.66 (2.32)	2.35–2.97	2.62 (2.71)	2.37–2.86
M	0.96 (2.12)	0.56–1.36	1.01 (2.13)	0.73–1.29	0.85 (1.94)	0.67–1.03
F	6.3 (4.8)	5.39–7.21	6.1 (4.44)	5.52–6.68	6.37 (4.57)	5.94–6.79
DMFT	9.68 (5.9)	8.56–10.79	9.77 (5.57)	9.03–10.5	9.83 (5.25)	9.35–10.31

*Abbreviations: CI – confidence interval, D – decayed teeth, DMFT – decayed, missing, repaired permanent teeth index, F – repaired/filled teeth, M – missing teeth, SD – standard deviation.

The results of the correlation analysis confirmed that the composition of the DMFT index was similar in all three samples (**Table 2**). The number of repaired teeth was most strongly correlated with the DMFT index, followed by a moderate correlation between the number of missing teeth (M) and the DMFT index and a low correlation between the number of decayed teeth and the DMFT index.

Table 2. Correlation matrices for the DMFT index and the components of the DMFT index in samples of the student population of the University of Split, presented as Pearson's correlation coefficient (R)*

	Small sample (n = 50)				Medium sample (n = 100)				Entire sample (n = 200)			
Variables	D	M	F	DMFT	D	M	F	DMFT	D	M	F	DMFT
D	1.00	-0.15	-0.20	0.17	1.00	-0.16	-0.16	0.23	1.00	-0.08	-0.28	0.25
M	-0.15	1.00	0.39	0.62	-0.16	1.00	0.34	0.58	-0.08	1.00	0.23	0.51
F	-0.20	0.39	1.00	0.87	-0.16	0.34	1.00	0.86	-0.28	0.23	1.00	0.80
DMFT	0.17	0.62	0.87	1.00	0.23	0.58	0.86	1.00	0.25	0.51	0.80	1.00

*Abbreviations: D – decayed teeth, DMFT – decayed, missing, repaired permanent teeth index, F – repaired/filled teeth, M – missing teeth.

Estimation of the mean DMFT index for student populations using the Monte Carlo method

For each of the three MCS procedures, the size of the student population was assumed to be n = 20 000. The DMFT index data for each of the three generated student populations included original data from the small, medium, and entire sample – specifically, we ran 400 iterations to generate the population from the small sample (MCS 50), 200 from the medium sample (MCS 100), and 200 from the entire sample (MCS 200). The estimated parameters of the DMFT index and the components of the DMFT index based on the data from the generated student populations of the University of Split are presented in **Table 3**.

The probabilities of the individual outcomes for the DMFT index in the student populations generated by the MCS were largely consistent with the probabilities measured in the original samples, except for the maximum scores for the DMFT index (**Figure 1**). The DMFT index in the original samples ranged from 0 to 27, and up to 37 (MCS 100 and MCS 200) or 38 (MCS 50) in the generated populations.

Table 3. Estimated parameters of DMFT index and DMFT index components for student populations of the University of Split generated by Monte Carlo method*

	MCS 50 (400 iterations)†		MCS 100 (200 iterations)‡		MCS 200 (100 iterations)§	
Parameters	Mean (SD)	99% CI	Mean (SD)	99% CI	Mean (SD)	99% CI
D	2.44 (2.3)	2.41–2.46	2.63 (2.95)	2.61–2.65	2.62 (2.77)	2.59–2.64
M	0.82 (1.92)	0.8–0.84	0.89 (2.01)	0.87–0.91	0.81 (1.89)	0.79–0.82
F	5.66 (4.35)	5.62–5.7	5.63 (4.16)	5.59–5.67	5.51 (4.24)	5.47–5.54
DMFT	8.92 (5.33)	8.87–8.97	9.15 (5.22)	9.09–9.19	8.93 (5.19)	8.88–8.98

* Abbreviations: CI – confidence interval, D – decayed teeth, DMFT – decayed, missing, repaired permanent teeth index, F – repaired/filled teeth, M – missing teeth, MCS – Monte Carlo simulation, SD – standard deviation.
† Population generated by Monte Carlo method from small sample.
‡ Population generated by Monte Carlo method from medium sample.
§ Population generated by Monte Carlo method from the entire sample.

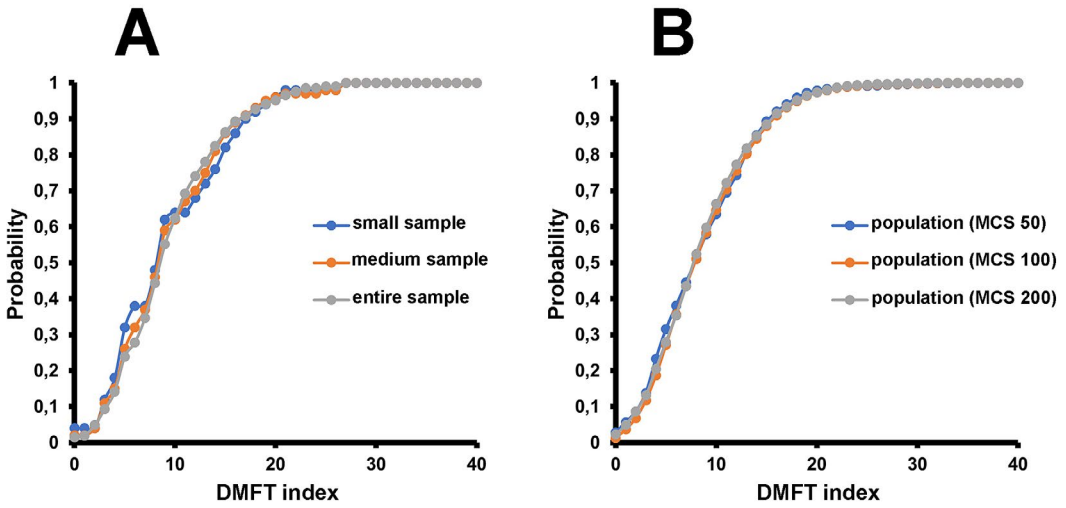

Figure 1. Probabilities of the individual outcomes for the DMFT index according to the cumulative distribution functions of the DMFT index in the original samples (A) and the student populations generated by the MCS (B) from the small (MCS 50), medium (MCS 100) and entire sample (MCS 200).

The results of the correlation analysis confirmed that the estimated composition of the DMFT index in all three student populations generated by MCS was like that of the original samples (**Table 4**). Among the components of the DMFT index, the number of repaired teeth was most strongly correlated with the DMFT index, with R values of 0.84 (MCS 50), 0.82 (MCS 100), and 0.77 (MCS 200).

The smaller the original sample on which the MCS was conducted, the greater the variability of the mean scores (**Figure 2**) and correlation coefficients (**Figure 3**) of the DMFT index and its components. Of all the variables, the DMFT index showed the least variability – only in some of the simulated samples were the mean scores for the DMFT index higher than the means measured in the corresponding original sample. Moreover, the mean scores of the DMFT index of the simulated samples in MCS 200 were not higher than the mean score measured in the original sample.

Table 4. Correlation matrices for the DMFT index and the components of the DMFT index in the student populations of the University of Split generated by Monte Carlo method, presented as Pearson's correlation coefficient (R)

	MCS 50†				MCS 100‡				MCS 200§			
Variables	D	M	F	DMFT	D	M	F	DMFT	D	M	F	DMFT
D	1.00	-0.06	-0.09	0.33	1.00	-0.06	-0.08	0.35	1.00	-0.02	-0.15	0.40
M	-0.06	1.00	0.18	0.48	-0.06	1.00	0.17	0.49	-0.02	1.00	0.10	0.43
F	-0.09	0.18	1.00	0.84	-0.08	0.17	1.00	0.82	-0.15	0.10	1.00	0.77
DMFT	0.33	0.48	0.84	1.00	0.35	0.49	0.82	1.00	0.40	0.43	0.77	1.00

*Abbreviations: D – decayed teeth, DMFT – decayed, missing, repaired permanent teeth index, F – repaired/filled teeth, M – missing teeth, MCS – Monte Carlo simulation.
† Population generated by Monte Carlo method from small sample.
‡ Population generated by Monte Carlo method from medium sample.
§ Population generated by Monte Carlo method from the entire sample.

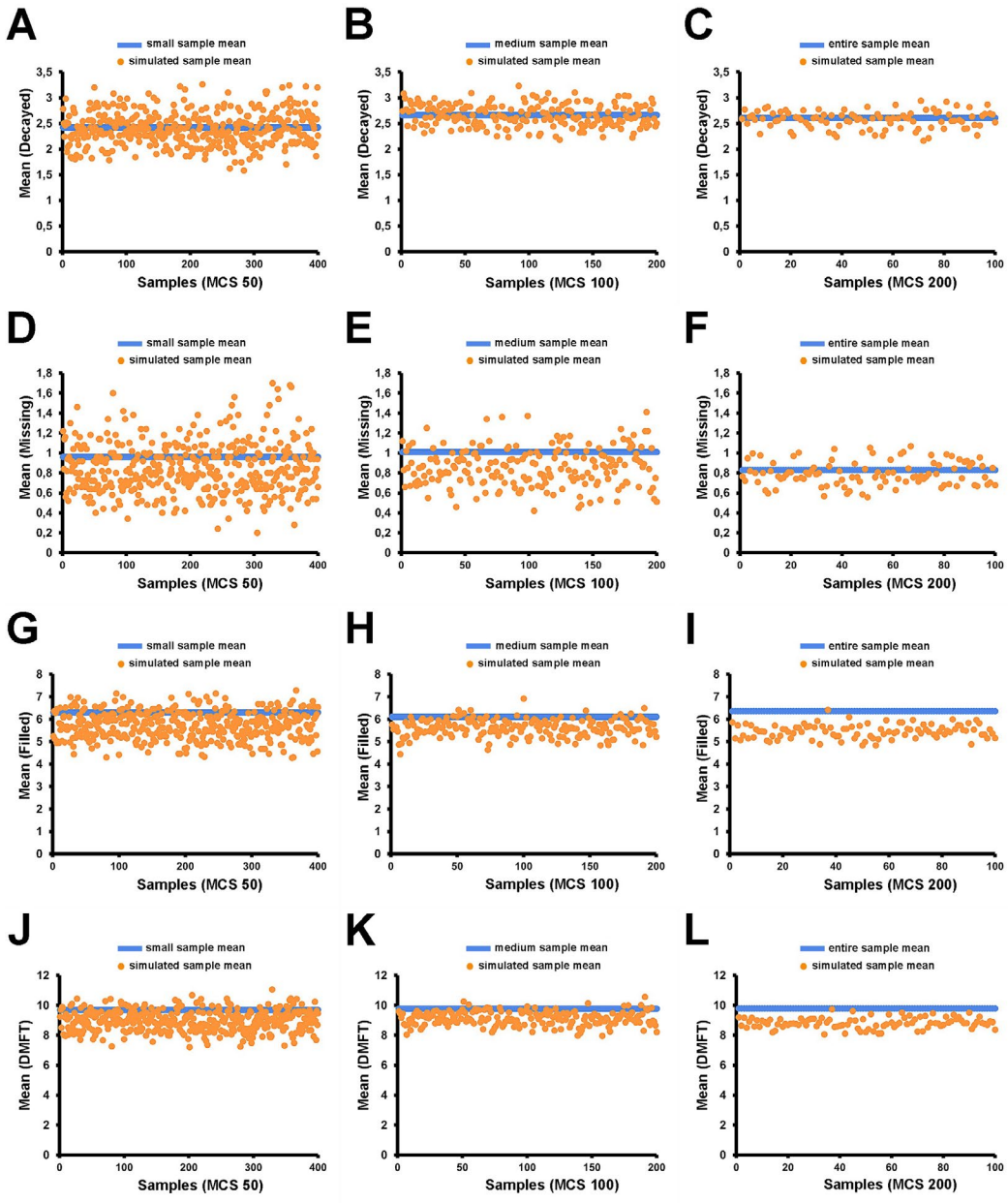

Figure 2. Variability of mean scores of DMFT index components (A-I) and DMFT index (J-L) in simulated samples derived by MCS from small (A, D, G, J), medium (B, E, H, K) and the entire (C, F, I, L) original samples. The blue horizontal lines show the mean scores measured in the original samples.

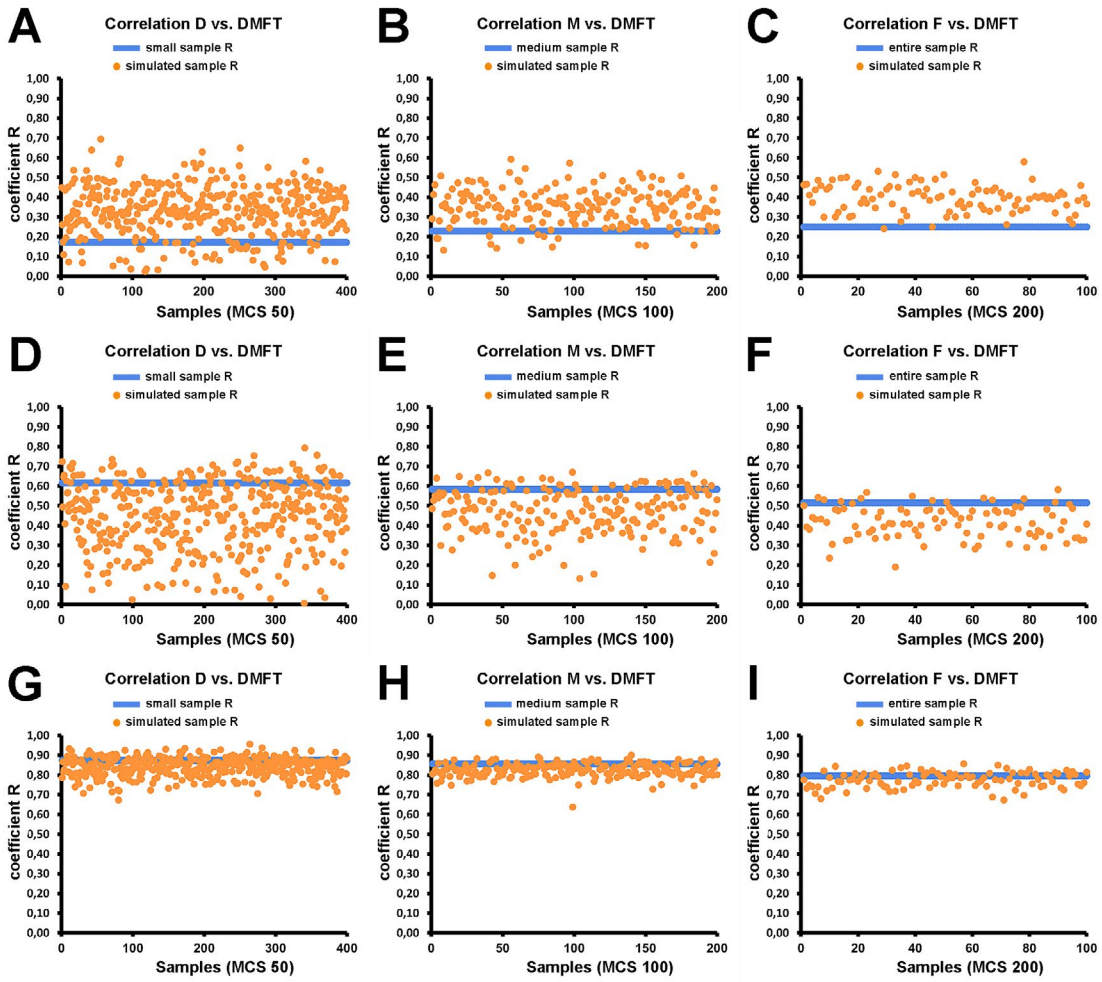

Figure 3. Variability of the correlation coefficient R in simulated samples derived with MCS from small (A, D, G), medium (B, E, H), and entire (C, F, I) original samples. The blue horizontal lines show the correlation coefficients measured in the original samples.

Estimating the mean DMFT index for student populations using the distribution of mean DMFT index scores from samples simulated with the Monte Carlo method

According to the Central Limit Theorem, the mean scores of the DMFT index of the simulated samples should be normally distributed, provided that the number of simulated samples is large enough. In addition, the central place in this distribution should belong to the actual mean of the DMFT index of the student population. The distributions of the mean scores of the DMFT index of all simulated samples are represented by separate curves for each MCS (**Figure 4**). The estimate for the mean DMFT index of the student population is 8.96 (SD = 0.69, 99% CI = 8.91–9.00) according to MCS 50, 9.13 (SD = 0.47, 99% CI = 9.08–9.17) according to MCS 100, and 8.83 (SD = 0.36, 99% CI = 8.78–8.87) according to MCS 200.

Of the three estimates of the mean DMFT index for the student population, the MCS 100 estimate was statistically significantly different from the MCS 50 and MCS 200 estimates, but this difference was less than one point (**Table 5**), which can be considered clinically negligible.

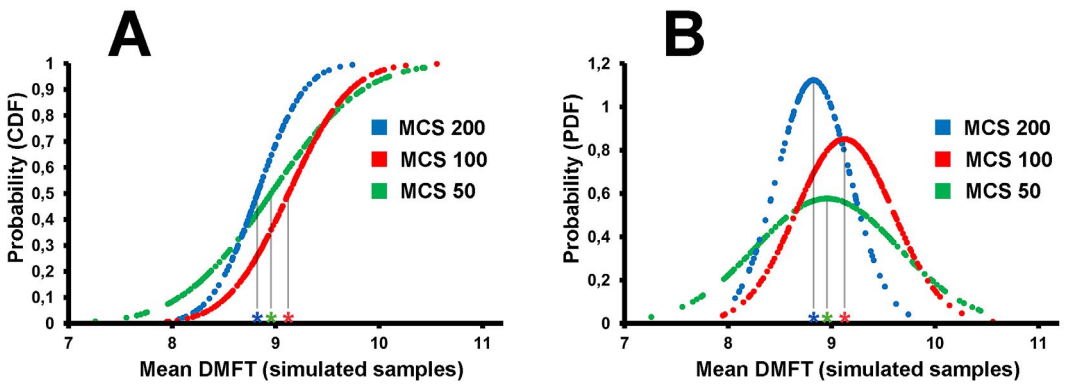

Figure 4. Distribution of the mean scores of the DMFT index in the simulated samples according to MCS 50 (green), MCS 100 (red) and MCS 200 (blue), represented by the cumulative distribution function (A) and the probability density function (B). The estimates for the actual mean DMFT index of the student population after each MCS are marked with asterisks.

Table 5. Comparison of the estimated means of the DMFT index for the student populations generated using the Monte Carlo method from small, medium and the entire original samples*

Outcome	'Dummy' variables (categories)	Predictors		Model parameters		
		Coefficient†	p-value	R	p-value‡	df
Mean DMFT¶	DMFT – MCS 200§	-	-	0.16	<0.001	2
	DMFT – MCS 100	0.13	0.049			
	DMFT – MCS 50	0.30	<0.001			

*Abbreviations: df – degrees of freedom, DMFT – decayed, missing, repaired permanent teeth index, MSC – Monte Carlo simulation.
† Expressed as the difference (positive/negative) to the average of the mean DMFT index scores of all simulated samples of MCS 200 as reference category.
‡ Multiple linear regression (α = 0.01 (*P* < 0.01).
§ Reference category ('dummy' coding), n = 100.
¶ Dependent variable – comprised of mean DMFT index scores of all simulated samples from MCS 50, MCS 100 and MCS 200 (n = 700).

Discussion

In this study, we generated data and assessed the mean score and composition of the DMFT index of the entire student population of the University of Split using the Monte Carlo method. We observed that the parameters of the DMFT index of the original samples and the simulated student populations determined by MCS (mean scores, SDs, and correlation coefficients) did not differ significantly. However, the CIs for the estimated mean scores were significantly narrower in the simulated student populations due to their size, as they were 100 to 400 times larger than the original sample.

Our estimates of the mean DMFT index of the target student population are largely consistent with those from similar studies previously conducted on samples of different sizes (from 50 to 1000 subjects) of adolescents and students from the University of Split and the University of Zagreb (*6*, *16*-*19*); their estimates ranged from 7 to 11 points for the DMFT index, which is within the variation range of the mean DMFT index for the samples simulated in our study using the Monte Carlo method.

The limitations of estimating the mean DMFT index using MCS, as we did in this study, arise from the fact that, although this estimate may be statistically justified, it might not correspond to the true population mean. The accuracy of the MCS is primarily determined by the distribution of the data in the sample and the entry criteria according to which the MCS is performed. The more similar the distribution of the data in the sample is to the actual distribution in the population (i.e., the more representative the sample is of the population), the more accurate the assessment of the parameters by MCS will be. The consistency of the data distribution in the sample and the population depends only partially on the sample size (*20*-*22*). For variables with a uniform distribution, precise parameter estimates based on small samples and a smaller number of iterations can be made when performing MCS. For variables with a normal distribution or an asymmetric/skewed distribution (e.g., exponential distribution), a larger sample is required for more accurate parameter estimation. Accordingly, MCS must be performed with a higher number of iterations to make the evaluation of the population means consistent with the postulates of the Central Limit Theorem (*1*). In this case, sample size matters because larger samples are more likely to be more representative of the population than small samples, meaning that the distribution of data in larger samples is more likely to resemble the actual distribution of data in a given population. Of course, this does not mean that representativeness cannot be achieved with a small sample, but it is just less likely (*12*, *23*, *24*).

The introduction of correlation between variables as an additional criterion for MCS is important because variables are rarely truly independent of each other. The DMFT index is a complex variable composed of the number of decayed, extracted, and repaired teeth; its components are partly interdependent. For example, the total number of decayed teeth may include teeth with secondary caries around fillings, while the missing/extracted teeth cannot be simultaneously counted as decayed or repaired teeth. Here we omitted the correlation between DMFT index and age, as the participants were in the same age group and as the overall age of the participants did not vary significantly. Accordingly, a larger initial sample size and the inclusion of more diverse age and other demographic variables could improve the accuracy on MCS when introducing their correlation with the DMFT index as an additional criterion for MCS. Likewise, we used convenience sampling in recruiting participants for this study, which is a non-probabilistic sampling method that may not provide a representative sample of the total population, thereby limiting the effectiveness of the MCS and the generalizability of the findings.

The Monte Carlo method can be useful in evaluating population parameters for clinical indices in dental medicine, such as the DMFT index. Its advantage is that the randomness of the parameters that occur in repeated samples can be simulated by computer in a short time. Moreover, MCS can be performed regardless of the sample size. However, a larger sample size is important when dealing with high variance variables and (consequently) when one wants to improve the precision of population parameter estimation using MCS by increasing either the number of variables and/or the criteria for MCS. Therefore, repeated sampling using probabilistic sampling methods and reassessment of clinical indices such as the DMFT index by combining the standard criteria for sample representativeness and the Monte Carlo method could be a good strategy to estimate the oral health status of a population.

Provenance: Submitted. This manuscript is based on the master's thesis by Dora Dodig, deposited in the Dabar repository (https://urn.nsk.hr/urn:nbn:hr:171:876527).

Received: 7 December 2023 / **Accepted:** 25 March 2024 / **Published online:** 24 April 2024.

Peer review: Externally peer-reviewed.

Ethics statement: The Ethics Committee of the University of Split School of Medicine approved the study (Class 003-08//23-03/0015; Reg. No. 2181-198-03-04-23-0082).

Availability of data: Raw data and Monte Carlo simulations used for the statistical analysis are available at the Open Science Framework: https://doi.org/10.17605/OSF.IO/H6ARV.

Funding: This research received no specific grant from any funding agency in the public, commercial or not-for-profit sectors.

Authorship declaration: DD participated in the collection of raw data, data analysis and interpretation of results, wrote the manuscript draft, and contributed to revisions of the manuscript. DK devised the concept and topic of the study, supervised, and performed data analysis and interpretation of results, wrote the manuscript, and contributed to revisions of the manuscript.

Disclosure of interest: The corresponding author completed the ICMJE Disclosure of Interest Form (available upon request from the corresponding author) and disclosed no relevant interests.

ORCID

Darko Kero https://orcid.org/0000-0002-8091-6347

References

1. Rudolph JE, Zhong Y, Duggal P, Mehta SH, Lau B. Defining representativeness of study samples in medical and population health research. BMJ Med. 2023;2(1):e000399. https://doi.org/10.1136/bmjmed-2022-000399
2. Gupta KK, Attri JP, Singh A, Kaur H, Kaur G. Basic concepts for sample size calculation: Critical step for any clinical trials! Saudi J Anaesth. 2016;10(3):328–31. https://doi.org/10.4103/1658-354X.174918
3. Serdar CC, Cihan M, Yucel D, Serdar MA. Sample size, power and effect size revisited: simplified and practical approaches in pre-clinical, clinical and laboratory studies. Biochem Med (Zagreb). 2021;31(1):010502. https://doi.org/10.11613/BM.2021.010502
4. Lim CY, In J. Randomization in clinical studies. Korean J Anesthesiol. 2019;72(3):221–32. https://doi.org/10.4097/kja.19049
5. Pincus T, Bergman MJ, Maclean R, Yazici Y. Complex measures and indices for clinical research compared with simple patient questionnaires to assess function, pain, and global estimates as rheumatology "vital signs" for usual clinical care. Rheum Dis Clin North Am. 2009;35(4):779–86. [ix.] https://doi.org/10.1016/j.rdc.2009.10.010
6. Cvitanović B. Procjena oralnog zdravlja studenata Medicinskog fakulteta Sveučilišta u Splitu temeljem KEP indeksa [master's thesis]. Split, Croatia: University of Split, School of Medicine; 2022.
7. Radić M, Benjak T, Vukres VD, Rotim Z, Zore IF. Presentation of DMFT/dmft Index in Croatia and Europe. Acta Stomatol Croat. 2015;49(4):275–84. https://doi.org/10.15644/asc49/4/2
8. Ambarkova V, Ivanova V. Dental caries experience among primary school children in the Eastern Region of the Republic of Macedonia. Oral Health Dent Manag. 2014;13(1):1–7.
9. Ambarkova V, Jankulovska M, Arian D, Glavina D, Soleva A. Dental caries experience among secondary school children in the Vardar region of the Republic of Macedonia. Oral Health Dent Manag. 2014;13(3):805–10.

10. Rothman KJ, Gallacher JE, Hatch EE. Why representativeness should be avoided. Int J Epidemiol. 2013;42(4):1012–4. https://doi.org/10.1093/ije/dys223

11. Taherdoost H. Sampling methods in research methodology; how to choose a sampling technique for research. International Journal of Academic Research in Management. 2016;5(2):18–27. https://doi.org/10.2139/ssrn.3205035

12. MIT OpenCorseWare. Introduction to computational thinking and data science – Confidence intervals [cited 2023 *18* Aug 2023]. Available from: https://www.youtube.com/watch?v=rUxP7TM8-wo&list=PLUl4u3cNGP619EG1wp0kT-7rDE_Az5TNd&index=8&t=9s.

13. Kroese DP, Brereton T, Taimre T, Botev ZI. Why the Monte Carlo method is so important today. Wiley Interdiscip Rev Comput Stat. 2014;6(6):386–92. https://doi.org/10.1002/wics.1314

14. Cohen J. Multiple regression as a general data-analytic system. Psychol Bull. 1968;70(6):426–43. https://doi.org/10.1037/h0026714

15. Delić I, Kero D. How much are the IMRaD structures of diploma theses and original research articles similar? A cross-sectional analysis of diploma theses from the Dental Medicine study in Split. ST-OPEN. 2023;4:e2023.2208.6. https://doi.org/10.48188/so.4.7

16. Ivica A, Galic N. Attitude towards Oral Health at Various Colleges of the University of Zagreb: A Pilot Study. Acta Stomatol Croat. 2014;48(2):140–6. https://doi.org/10.15644/asc48/2.140

17. Jurković A. Epidemiološka obilježja karijesa kod studenata Stomatološkog fakulteta u Zagrebu [master's thesis]. Zagreb, Croatia: University of Zagreb, School of Dental Medicine; 2018.

18. Tepić A. Oralno zdravlje stomatologa [master's thesis]. Zagreb, Croatia: University of Zagreb, School of Dental Medicine; 2017.

19. Vranjković H. Radiografska procjena dentalnog statusa adolescenata Splitsko-dalmatinske županije [master's thesis]. Split, Croatia: University of Split School of Medicine; 2016.

20. Hannigan A, Lynch CD. Statistical methodology in oral and dental research: pitfalls and recommendations. J Dent. 2013;41(5):385–92. https://doi.org/10.1016/j.jdent.2013.02.013

21. Pandis N, Polychronopoulou A, Eliades T. An assessment of quality characteristics of randomised control trials published in dental journals. J Dent. 2010;38(9):713–21. https://doi.org/10.1016/j.jdent.2010.05.014

22. Yan F, Robert M, Li Y. Statistical methods and common problems in medical or biomedical science research. Int J Physiol Pathophysiol Pharmacol. 2017;9(5):157–63.

23. Limpert E, Stahel WA. Problems with using the normal distribution--and ways to improve quality and efficiency of data analysis. PLoS One. 2011;6(7):e21403. https://doi.org/10.1371/journal.pone.0021403

24. Szucs D. A Tutorial on Hunting Statistical Significance by Chasing N. Front Psychol. 2016;7:1444. https://doi.org/10.3389/fpsyg.2016.01444