


BMJ Open Identifying heterogeneity in the risk factors of dental caries status in Chinese adolescents using Poisson mixture regression

Chao Yuan,¹ Jie He,² Xiangyu Sun,¹ Jian Kang,² Shuguo Zheng ¹

To cite: Yuan C, He J, Sun X, *et al.* Identifying heterogeneity in the risk factors of dental caries status in Chinese adolescents using Poisson mixture regression. *BMJ Open* 2020;**10**:e039599. doi:10.1136/bmjopen-2020-039599

► Prepublication history and additional materials for this paper is available online. To view these files, please visit the journal online (<http://dx.doi.org/10.1136/bmjopen-2020-039599>).

CY and JH contributed equally.

Received 20 April 2020

Revised 12 September 2020

Accepted 22 September 2020



© Author(s) (or their employer(s)) 2020. Re-use permitted under CC BY-NC. No commercial re-use. See rights and permissions. Published by BMJ.

¹Department of Preventive Dentistry, Peking University School and Hospital of Stomatology, National Engineering Laboratory for Digital and Material Technology of Stomatology, Beijing Key Laboratory of Digital Stomatology, Beijing, People's Republic of China

²Department of Biostatistics, University of Michigan, Ann Arbor, Michigan, USA

Correspondence to

Dr Shuguo Zheng; kqzsg86@bjmu.edu.cn and Dr Jian Kang; jiankang@umich.edu

ABSTRACT

Objective The purpose of this study was to cluster individuals into groups with different dental health characteristics and make statistical inferences on the effect differences among different groups simultaneously to identify the heterogeneity of risk factors in Chinese adolescents by analysing the data from the 4th Chinese National Oral Health Survey.

Methods For decayed, missing and filled permanent teeth (DMFT), mean values were statistically analysed for their relationships with different categories of all involved variables. As DMFT scores only have discrete values, Poisson mixture regression was adopted to model the heterogeneity and complex patterns in the association and to detect the subgroup. The Bayesian information criterion (BIC) was used to determine the optimal number of subgroups. A series of Wald tests were used to explore the relationship between risk factors including the interaction effects and the number of DMFT.

Results A total of 100 986 individuals aged 12–15 years old were analysed. The model clustered different individuals into three subgroups and built three submodels for detailed statistical inference simultaneously. The number of individuals in the three subgroups were 52 576 (52.1%), 41 969 (41.5%) and 6441 (6.4%), respectively. The mean (SD) of DMFT of the three subgroups was 0.50 (1.05), 0.99 (1.21), 5.59 (2.50). The model fitting results indicated that the effects of all risk factors on DMFT appear to be different in three subgroups. Controlling the confounding effects, our analysis implied that the regional inequality was the main contributing factor to dental caries among adolescents in Chinese mainland.

Conclusions The risk factors of dental caries exhibited heterogeneity in groups with different characteristics. The Poisson mixture regression model could cluster individuals into groups and identify the heterogeneous effects of risk factors among different groups. The findings support the need for different targeted interventions and prevention measures in groups with different dental health characteristics.

INTRODUCTION

The adolescent period is an important stage of children's growth and development.¹ After the replacement of primary teeth by their successors, new dentition—namely,

Strengths and limitations of this study

- Data from the 4th Chinese National Oral Health Survey were analysed to obtain the prevalence and associated factors of permanent dental caries in Chinese adolescents aged 12–15 years old.
- It is advantageous for epidemiological surveys to cluster individuals into groups with different dental health characteristics and therefore make targeted statistical inferences to identify the heterogeneity of risk factors.
- Poisson mixture regression is a sound statistical model for the subgroup analyses. It has the flexibility in modelling the zero-inflated count measure of dental decay and characterising the heterogeneous risk factors of dental caries.
- The modified Expectation Maximization (EM) algorithm is an efficient parameter estimation method for Poisson mixture regression, while the Bayesian information criterion can be used to determine the number of subgroups.
- This research on dental caries using cross-sectional data collection could not represent casual relationships between dental caries and the related factors, and hence, was unable to indicate how caries changes over time in the general population or in a given sample and what variables are associated with the overall trend and the differences within and among persons.

permanent dentition—commences and functions during the following decades of life. Thus, the WHO regards 12 years old to 15 years old as a reference group for population oral health surveys.² Dental caries of permanent teeth is the most prevalent oral disease and, according to the Global Burden of Disease Study 2016, is the disease with the second highest incidence.³

During the last two decades, the prevalence of permanent caries in 12-year-old Chinese children fluctuated, with rates of 45.8%, 28.9% and 38.5% in 1995, 2005 and 2015, respectively.^{4–6} Meanwhile, the characteristics of the Chinese

population are steadily changing, with a decrease in the size of the rural population from 57.0% in 2005 to 43.9% in 2015. As the inequality of economic development, the lifestyles of the adolescents' families were quite different. The adolescents' oral disease patterns are expected to appear the heterogeneity. In particular, the risk factors for dental caries can be different for different groups characterised by their individual characteristics such as demographics and social economic status. For example, the geographic heterogeneity analysis of dental caries demonstrated inequalities in decayed, missing and filled permanent teeth (DMFT) among different regions of the country.⁷ The oral behaviours of adolescents, including toothbrushing habits, consumption of sugary foods and dental attendance, also vary in different provinces and urban/rural areas.^{8,9} This motivates the needs of clustering individuals into groups in light of the heterogenous associations between the individual characteristics and the risk of dental caries. It is advantageous for epidemiological surveys to cluster individuals into groups with different dental health characteristics and therefore make targeted statistical inferences to identify the heterogeneity of risk factors.

In this work, we propose to use the Poisson mixture regression model to analyse the dental dataset and to perform the subgroup detection to identify the heterogeneity of risk factors. The finite mixture regression model was first proposed by Wolfe,¹⁰ and has been successfully applied to solve many problems in various research areas.^{11,12} The Poisson mixture regression model is a systematic statistical approach that explicitly models the mixture distribution of the regression coefficients, leading to clear interpretation of fitting results for the subgroup analyses. There is a formal statistical procedure to determine the number of mixture components in the model which corresponds to the number of subgroups in the population. A suitable probability distribution model can also be added into the model according to the characteristics of the outcome variables of interest, resulting in a more flexible and adaptive data analysis procedure.^{13,14}

As dental caries are posing great disease burdens to the adolescent population,¹⁵ caries control has been put onto the agenda of the government by carrying out national public health programmes with broad application of appropriate preventive techniques for children. Based on all of the above, the aim of this study is to use the Poisson mixture regression model to cluster individuals into groups and to identify the heterogeneity of risk factors for dental caries in Chinese adolescents. Analysis of the most recent prevalence data of caries among adolescents and its related factors will yield better, more reasonable suggestions on the reformation and innovation of policies related to dental public health.

MATERIALS AND METHODS

Description of data

This cross-sectional study used data from the 4th National Oral Health Survey of China (2015–2016). All 31 provinces

of the China Mainland participated in the survey. The specific survey description is supplied in online supplemental appendix. Data of the 12–15 years old age group was used for the present study.

Poststratification weights were used to adjust for differences in the age-by-sex-by-location-by-province distribution between the sample and the general population in the 31 provinces involved in the study, consistent with the 6th National Demographic Census in 2010.^{16,17}

The number of DMFT was calculated to determine the overall caries status and experience. The risk factors involved in the present study included six categories: social demographic factors, oral hygiene behaviours, sugar consumption habits, utilisation of dental service, oral health knowledge and pit-and-fissure sealants. Detailed information on the grading standard of each variable is presented in online supplemental appendix table 1. The variables extracted from the examination tables and questionnaires were representative oral health-related factors, which were further statistically analysed with regard to their relationships with dental caries experience.

Patient and public involvement

All participants were selected using multistage stratified cluster sampling method and provided written informed consent to take part in the study.

Model and estimation

Let y denote the response variable, that is, the number of DMFT, taking the count values from 0 to 28. Let x represent a p dimensional covariate vector including all the risk factors and their interactions as well as other confounding factors. We consider a Poisson mixture regression model, where the conditional distribution of y given x is specified as follows:

$$H(y|x, \omega; \Theta) = \sum_{k=1}^K \pi_k \text{Poi}(y|\lambda_k(x, \beta^{(k)})),$$

$$\log \lambda_k(x, \beta^{(k)}) = \sum_{j=1}^p \beta_j^{(k)} x_j, \quad k = 1, \dots, K, \quad (1)$$

where K is the number of subgroups, $x = (x_1, \dots, x_p)^T$ and $\beta^{(k)} = (\beta_1^{(k)}, \dots, \beta_p^{(k)})^T$. For the k th subgroup, $\text{Poi}(y|\lambda_k(x, \beta^{(k)})) = \frac{\lambda_k(x, \beta^{(k)})^y}{y!} e^{-\lambda_k(x, \beta^{(k)})}$ is a Poisson probability mass function with mean intensity $\lambda_k(x, \beta^{(k)})$, and $\beta^{(k)}$ represents the effects of x . $\{\pi_k\}_{k=1}^K$ are probability weights that are associated with component distributions $[\text{Poi}(y|\lambda_k(x, \beta^{(k)}))]_{k=1}^K$, which satisfy

$$0 \leq \pi_k \leq 1, \quad k = 1, \dots, K \quad \text{and} \quad \sum_{k=1}^K \pi_k = 1$$

In model (1), $\beta_j^{(k)}$ represents the effect of covariate x_j in subgroup k . When the values of other covariates $x_{-j} = (x_1, \dots, x_{j-1}, x_{j+1}, \dots, x_p)^T$ are fixed, from model (1), we have

$$\beta_j^{(k)}(x_j^{(1)} - x_j^{(2)}) = \log[\lambda_k(x_{-j}, x_j^{(1)})] - \log[\lambda_k(x_{-j}, x_j^{(2)})],$$

where $x_j^{(1)}$, $x_j^{(2)}$ represent two different values of x_j . When $x_j^{(1)} - x_j^{(2)} = 1$,

$$\beta_j^{(k)} = \log[\lambda_k(x_{-j}, x_j^{(1)})] - \log[\lambda_k(x_{-j}, x_j^{(2)})],$$

which is the change of the log mean of y in subgroup k per unit increase in x_j , defined as the effect size of x_j . When $\beta_j^{(k)} > 0$, x_j has positive effects on the average of y . Similarly, when $\beta_j^{(k)} < 0$, the effect of x_j on y goes in the opposite direction. In addition, by comparing $\beta_j^{(k)}$, $k = 1, \dots, K$ with fixed j , we also could infer the effects of factor x_j on the response among different subgroups. We summarise the estimation procedure of β in online supplemental appendix.

All of the above analyses were performed with the R package ‘FlexMix’.^{18–20} As DMFT scores only have discrete values, we employed the Poisson mixture regression model into the procedure of subgroup association detection. An optimised group number was selected from a pre-specified number established based on the Bayesian information criterion (BIC)^{21 22} with all computation completed by the modified EM algorithm. In each iteration of the M-step, the algorithm will remove the subgroups with the probability weight below a given threshold set as 5%²⁰ for meaningful interpretations of subgroups, ensuring that the estimated subgroups all have a sample size large than 5% of total. We defined that the estimates represented the effect of the risk factors on DMFT in each subgroup. The estimates were compared in detail among the different subgroups. A series of Wald tests were also conducted to explore the relationship between the number of DMFT and risk factors with an interaction effect. $P < 0.05$ was regarded as statistical significance (two-sided).

RESULTS

The data set consisted of 118601 individuals aged 12–15 years old. We focused on 100986 individuals with complete observations. The model fitting results with different pre-specified number of subgroups (k_0) are summarised in online supplemental appendix table 2. The estimated group number was denoted as k ($k \leq k_0$). In our analysis, the k values were always equal to 3 under circumstances of $3 \leq k_0 \leq 6$. By comparisons of the BIC, we conclude that three subgroups could sufficiently represent the heterogeneity of risk factors of dental caries among the 12 years old population.

In the three subgroups (subgroup 1–3), the number of individuals were 52576, 41969 and 6441. The estimated proportions of three subgroups were $\hat{\pi}_1 = 0.52$, $\hat{\pi}_2 = 0.42$ and $\hat{\pi}_3 = 0.06$, respectively. The mean DMFT values (SD) of the three subgroups were 0.50 (1.05), 0.99 (1.21) and 5.59 (2.50), which had statistical difference among subgroups ($p < 0.001$). In subgroup 1, 50.8% were boys, while 49.2% were girls. In subgroup 2, there were 48.6% boys and 51.4% girls. In subgroup 3, 50.4% were boys and 49.6% were girls. The study population was also evenly distributed to urban (37.1% in subgroup 1, 42.5% in subgroup 2 and 39.7% in subgroup 3) and rural (62.9% in subgroup 1, 57.5% in subgroup 2 and 60.3% in subgroup 3) areas. In the three subgroups (subgroup 1–3), 86.4%, 87.6% and 86.1% children were reported to brush their

teeth for more than once a day. The summary information of subgrouping results are shown in table 1. The marginal distribution of observed and fitted DMFT within each subgroup are summarised in online supplemental appendix table 1 and figure 1. The fitted R-squared value for this model is 0.785.

Table 2 shows the parameter estimates and the 95% CIs of the three subgroups in the model. The results of whether a significant difference existed among the effects of the same variable in the three subgroups are summarised in online supplemental appendix tables 4 and 5.

Social demographic factors

The effects of regional factors in the eastern area were strongest in subgroups 1 and 2, but the influence sharply decreased with the increasing severity of DMFT in subgroup 3 ($p < 0.001$). Adolescents in the eastern region had the largest DMFT in subgroup 1 and the smallest DMFT in subgroup 2. However, the differences among the different subgroups in the western and central regions were not significant ($p > 0.05$). The adolescents with older age were more likely to have more DMFT. The influence on subgroup 3 was the strongest ($p < 0.05$). The average DMFT of males was less than that of females in all subgroups. Adolescents in urban areas had fewer DMFT in all subgroups, but the influence decreased with the increasing severity of DMFT. The estimates of Only child and educational level factor were less than 0.1 in all subgroups.

Oral hygiene behaviour

Adolescents who brushed their teeth every day had fewer DMFT in all subgroups, but the influence decreased with the increasing severity of DMFT. The estimates were -1.606 , -1.231 and -0.849 in subgroups 1, 2 and 3, respectively.

Sugar consumption habits

Adolescents with a higher frequency of sugar consumption had more DMFT in all subgroups, but the influence decreased as the severity of DMFT increased.

Pit-and-fissure sealant history

Adolescents who had received pit-and-fissure sealant had fewer DMFT in all subgroups, but the influence decreased with the increasing severity of DMFT in subgroup 3 ($p < 0.001$). The estimates were -0.178 , -0.129 and -0.095 in subgroups 1, 2 and 3, respectively.

Oral health knowledge

The higher the adolescents’ oral health knowledge, the fewer DMFT in all subgroups. However, all estimates were less than -0.01 .

Interaction effects

Tables 3 and 4 summarise the estimated effects for region factor X_1 and X_2 , under the different levels of the sugar intake factor X_{10} . With increasing sugar consumption,

Table 1 The summary information of the three subgroups' results

Categories	Subcategories	Subgroup 1	Subgroup 2	Subgroup 3	Total
N		52 576	41 969	6 441	100 986
Mean of DMFT (SD)		0.50 (1.05)	0.99 (1.21)	5.59 (2.50)	1.04 (1.76)
Mean of age (SD)		14.02 (1.10)	14.02 (1.09)	14.10 (1.09)	14.03 (1.09)
Mean of score of sugar consumption (SD)		7.69 (3.51)	7.97 (3.49)	7.86 (3.52)	7.81 (3.51)
Mean of score of oral knowledge (SD)		4.81 (1.82)	4.92 (1.85)	4.86 (1.90)	4.86 (1.83)
Region 1 (%)	East	24.6	47.4	38.2	35.0
	Otherwise	75.4	52.6	61.8	65.0
Region 2 (%)	West	45.8	32.6	38.7	39.9
	Otherwise	54.2	67.4	61.3	60.1
Gender (%)	Male	50.8	48.6	50.4	50.1
	Female	49.2	51.4	49.6	49.9
Type of census register (%)	Urban	37.1	42.5	39.7	39.5
	Rural	62.9	57.5	60.3	60.5
Only child (%)		33.1	39.5	36.7	36.0
Paternal educational level (%)	Never accepted education	2.1	2.5	3.4	2.4
	Elementary school	19.4	17.3	18.6	18.5
	Middle school	46.8	43.8	42.7	45.3
	High school	16.8	17.3	16.6	17.0
	Secondary school	3.9	4.9	4.6	4.4
	College	5.0	6.5	6.8	5.8
	Undergraduate	5.0	6.5	6.2	5.7
Maternal educational level (%)	Graduate or higher	0.8	1.2	1.0	1.0
	Never accepted education	6.6	6.2	7.4	6.5
	Elementary school	26.0	22.3	23.0	24.3
	Middle school	40.8	40.0	39.8	40.4
	High school	13.1	13.6	13.1	13.3
	Secondary school	4.4	5.6	5.1	4.9
	College	4.4	6.0	5.7	5.2
Frequency of toothbrushing (%)	Undergraduate	4.2	5.4	5.0	4.7
	Graduate or higher	0.6	0.9	0.9	0.7
	Daily or often	86.4	87.6	86.1	86.8
	Seldom or never	13.6	12.4	13.9	13.2
Pit-and-fissure sealant history (%)		4.8	6.9	5.9	5.7
Self-assessment to tooth (%)	Very good	5.6	6.2	7.2	6.0
	Good	28.5	29.8	28.0	29.0
	Fair	52.8	50.5	46.8	51.5
	Poor	11.0	11.3	14.7	11.3
	Bad	2.1	2.2	3.2	2.2
Visit dentist (%)	Yes	47.2	51.9	52.8	49.5
	No	52.8	48.1	47.2	50.5

DMFT, decayed, missing and filled permanent teeth.

the effects of the eastern regional factor gradually weakened, while the effects of the central and western regions became significant and gradually increased. Table 5 summarises the estimated effects of X_{10} under different values of X_1 and X_2 . In subgroup 1, the effect of sugar consumption was greatest in the western region (0.130),

second in the central region (0.100) and lowest in the eastern region (0.019). In subgroup 2, the influence of sugar consumption was highest in the eastern region (0.122), second highest in the western region (0.051) and lowest in the central region (0.028). In subgroup 3, the influence of sugar consumption was greatest in the

Table 2 Estimation and 95% CI of three different subgroups in model (1)

	Subgroup 1			Subgroup 2			Subgroup 3		
	Estimation	SE	CI	Estimation	SE	CI	Estimation	SE	CI
Intercept	-4.802	0.090	-4.258 to 3.906	-0.763	0.044	-0.850 to 0.677	-0.215	0.039	-0.291 to 0.139
X ₁ (region 1)	3.464	0.076	3.315 to 3.613	-3.949	0.133	-4.210 to 3.688	0.304	0.018	0.269 to 0.339
X ₂ (region 2)	-0.035	0.085	-0.202 to 0.132	-0.004	0.016	-0.035 to 0.027	0.108	0.017	0.075 to 0.141
X ₃ (age)	0.083	0.004	0.075 to 0.091	0.133	0.003	0.127 to 0.139	0.161	0.003	0.155 to 0.167
X ₄ (gender)	-0.392	0.008	-0.408 to 0.376	-0.390	0.007	-0.404, to 0.376	-0.203	0.006	-0.215 to 0.191
X ₅ (urban/rural)	-0.124	0.009	-0.142 to 0.106	-0.034	0.008	-0.050 to 0.018	-0.019	0.007	-0.033 to 0.005
X ₆ (only child)	-0.090	0.009	-0.108 to 0.072	0.090	0.008	0.074 to 0.106	0.028	0.006	0.016 to 0.039
X ₇ (education 1)	-0.035	0.004	-0.043 to 0.027	-0.060	0.003	-0.066, to 0.054	-0.019	0.003	-0.025 to 0.013
X ₈ (education 2)	-0.010	0.004	-0.018 to 0.002	-0.009	0.003	-0.015 to 0.003	-0.014	0.003	-0.020 to 0.008
X ₉ (toothbrushing)	-1.606	0.021	-1.647 to 1.565	-1.231	0.017	-1.264 to 1.198	-0.849	0.013	-0.874 to 0.824
X ₁₀ (sugar)	0.100	0.007	0.086 to 0.114	0.028	0.001	0.026 to 0.030	0.019	0.002	0.015 to 0.023
X ₁₁ (sealant)	-0.178	0.014	-0.205 to 0.151	-0.129	0.018	-0.164 to 0.094	-0.095	0.013	-0.120 to 0.070
X ₁₂ (knowledge)	-0.008	0.002	-0.012 to 0.004	-0.005	0.002	-0.009 to 0.001	-0.004	0.001	-0.006 to 0.002

'Estimation' is the estimation of effect parameters and 'CI' is the 95% CI of parameter estimation.

**Table 3** Estimation and 95% CI for Wald test to the effect of regional factors X_1 under the case of controlling sugar intake factor X_{10} at different levels in different subgroups

	Subgroup 1			Subgroup 2			Subgroup 3		
	Coef.	SE	CI	Coef.	SE	CI	Coef.	SE	CI
$X_1 X_{10}$	-0.081	0.007	-0.095 to 0.067	0.094	0.011	0.072 to 0.116	-0.013	0.002	-0.017 to 0.009
$X_{10}(1)$	3.382	0.070	3.031 to 3.733	-3.855	0.123	-4.471 to 3.239	0.291	0.016	0.212 to 0.370
$X_{10}(2)$	3.301	0.064	2.980 to 3.622	-3.761	0.113	-4.328 to 3.194	0.278	0.014	0.208 to 0.349
$X_{10}(3)$	3.220	0.058	2.927 to 3.512	-3.668	0.103	-4.186 to 3.149	0.266	0.012	0.204 to 0.328
$X_{10}(4)$	3.138	0.053	2.874 to 3.403	-3.574	0.094	-4.046 to 3.102	0.253	0.011	0.199 to 0.308
$X_{10}(5)$	3.057	0.048	2.818 to 3.296	-3.480	0.085	-3.907 to 3.054	0.241	0.010	0.194 to 0.288
$X_{10}(6)$	2.976	0.043	2.761 to 3.191	-3.386	0.076	-3.770 to 3.003	0.228	0.008	0.187 to 0.270
$X_{10}(7)$	2.895	0.039	2.700 to 3.089	-3.293	0.068	-3.636 to 2.950	0.216	0.007	0.179 to 0.253
$X_{10}(8)$	2.813	0.036	2.635 to 2.992	-3.199	0.062	-3.506 to 2.892	0.203	0.007	0.168 to 0.239
$X_{10}(9)$	2.732	0.033	2.564 to 2.899	-3.105	0.550	-3.382 to 2.829	0.191	0.007	0.154 to 0.228
$X_{10}(10)$	2.651	0.032	2.487 to 2.814	-3.012	0.506	-3.266 to 2.758	0.178	0.008	0.138 to 0.219
$X_{10}(11)$	2.569	0.033	2.403 to 2.735	-2.918	0.048	-3.160 to 2.677	0.166	0.009	0.120 to 0.211
$X_{10}(12)$	2.488	0.035	2.313 to 2.663	-2.824	0.048	-3.066 to 2.584	0.153	0.010	0.101 to 0.206
$X_{10}(13)$	2.407	0.038	2.216 to 2.597	-2.731	0.050	-2.982 to 2.479	0.141	0.012	0.081 to 0.201
$X_{10}(14)$	2.325	0.042	2.115 to 2.535	-2.637	0.054	-2.910 to 2.364	0.128	0.014	0.060 to 0.196
$X_{10}(15)$	2.244	0.046	2.011 to 2.477	-2.543	0.060	-2.846 to 2.241	0.116	0.015	0.039 to 0.193
$X_{10}(16)$	2.163	0.051	1.904 to 2.421	-2.450	0.067	-2.787 to 2.112	0.103	0.017	0.018 to 0.189

X_{10} stands for the value of factor X_{10} . 'Coef.' is the estimation of the effect of X_1 under different cases, 'SE' stands for the SE of Wald statistics in testing Coef. is 0 or not.

western region (0.030), second in the central region (0.019) and lowest in the eastern region (0.007). With an increase in DMFT, the influence of the sugar consumption factor in the eastern and central regions gradually weakened. Table 6 shows the estimated effects of X_9 , the frequency of tooth brushing, under different levels of utilisation of the dental service factors X_{13} and X_{14} . Adolescents who brushed their teeth every day had fewer DMFT when their self-assessment of their teeth was 'very good', 'good' or 'fair'. However, as the self-assessment decreased, the influence gradually weakened until the effect reversed. The same phenomenon occurred for the associated factor of dental experience. The influence of toothbrushing frequency on reduced DMFT was weakened when adolescents had dental experiences.

DISCUSSION

This study used data sourced from the 4th National Oral Health Survey of China, which was conducted in 2015–2016 and covered all 31 provinces, municipalities and autonomous regions in the China Mainland. In the analysis in our previous study,⁶ we obtained the prevalence and associated factors of permanent dental caries in Chinese adolescents aged 12–15 years old and presented a descriptive analysis of the current condition of dental caries.

The mixture regression model is a novel approach to perform subgroup analysis and is applicable to many different types of data, both discrete and continuous. In contrast to the original regression method, the mixture regression model method partitions all individuals into data-driven subgroups and builds models with different coefficients for different subgroups. Rather than discovering the relationships among factors of interest and oral disease patterns by assuming all individuals share common effects, as in general regression analysis, we can make further comparisons of the effects of factors on the status of dental caries among different subgroups. This method provides more detailed and accurate results because the influence of the same factor may be completely different in different populations. In doing so, we can easily find commonalities as well as differences among different groups. Thus, the results of the analysis are more informative, and we can provide more detailed and helpful suggestions to populations according to their characteristics. In reality, all effects can be clustered into three different types: constant, varying and nested varying effects, which were varying in the first step and classified by Wald tests based on the model fitting results. In the second step, we were able to refit the model by adding classifications of all effects and obtain more accurate estimation results.

Table 4 Estimation and 95% CI for Wald test to the effect of regional factors X_2 under the case of controlling sugar intake factor X_{10} at different levels in different subgroups

	Subgroup 1			Subgroup 2			Subgroup 3		
	Coef.	SE	CI	Coef.	SE	CI	Coef.	SE	CI
X_2X_{10}	0.030	0.008	0.014 to 0.046	0.023	0.002	0.019 to 0.027	0.011	0.002	0.007 to 0.149
$X_{10}(1)$	-0.006	0.077	-0.394 to 0.383	0.018	0.014	-0.054 to 0.090	0.119	0.016	0.040 to 0.197
$X_{10}(2)$	0.024	0.070	-0.326 to 0.375	0.041	0.013	-0.023 to 0.105	0.129	0.014	0.060 to 0.199
$X_{10}(3)$	0.054	0.062	-0.259 to 0.367	0.063	0.011	0.007 to 0.120	0.140	0.012	0.079 to 0.202
$X_{10}(4)$	0.083	0.055	-0.193 to 0.360	0.086	0.010	0.037 to 0.135	0.151	0.011	0.097 to 0.205
$X_{10}(5)$	0.113	0.048	-0.129 to 0.356	0.109	0.009	0.066 to 0.151	0.162	0.009	0.115 to 0.209
$X_{10}(6)$	0.142	0.042	-0.067 to 0.353	0.131	0.007	0.094 to 0.169	0.172	0.008	0.131 to 0.214
$X_{10}(7)$	0.173	0.036	-0.009 to 0.354	0.154	0.007	0.120 to 0.187	0.183	0.007	0.146 to 0.221
$X_{10}(8)$	0.202	0.032	0.043 to 0.362	0.176	0.006	0.145 to 0.208	0.194	0.007	0.158 to 0.230
$X_{10}(9)$	0.231	0.029	0.086 to 0.377	0.199	0.006	0.166 to 0.231	0.205	0.007	0.168 to 0.242
$X_{10}(10)$	0.262	0.028	0.119 to 0.405	0.221	0.007	0.186 to 0.257	0.216	0.008	0.174 to 0.257
$X_{10}(11)$	0.291	0.030	0.139 to 0.443	0.244	0.008	0.203 to 0.284	0.226	0.009	0.180 to 0.273
$X_{10}(12)$	0.321	0.034	0.150 to 0.492	0.267	0.009	0.220 to 0.313	0.237	0.011	0.184 to 0.291
$X_{10}(13)$	0.351	0.039	0.154 to 0.548	0.289	0.011	0.236 to 0.343	0.248	0.012	0.187 to 0.309
$X_{10}(14)$	0.380	0.045	0.153 to 0.608	0.312	0.012	0.251 to 0.373	0.258	0.014	0.189 to 0.328
$X_{10}(15)$	0.410	0.052	0.149 to 0.671	0.334	0.014	0.265 to 0.404	0.269	0.016	0.191 to 0.347
$X_{10}(16)$	0.440	0.059	0.143 to 0.737	0.357	0.015	0.280 to 0.434	0.280	0.017	0.193 to 0.367

X_{10} stands for the value of factor X_{10} , other notations are of the same with [table 2](#).

According to model (1), the conditional distribution of DMFT for an individual i in subgroup k given x_i was specified as a Poisson distribution with mean $\exp(x_i\beta^{(k)})$, a function of covariate x_i . Thus, the conditional distribution was individual-specific and the marginal distribution of DMFT across individuals did not simply follow a Poisson distribution. The fitted R-squared value of Poisson mixture regression was 0.785, and the marginal distribution of fitted DMFT was very close to that of the observed DMFT within each subgroup (see online supplemental appendix table 3 and figure 1). We also evaluated the model fitting of Poisson regression and zero-inflated Poisson (ZIP) regression. The fitted results were summarised in online supplemental appendix figure 2, tables 3 and 6. The results show that for both ZIP and Poisson regression models, the fitted zeros are much smaller than the

observed zeros, while the fitted ones are much larger than the observed ones. This implies that both models do NOT fit the data very well, although the ZIP model may slightly improve the fitted number of zeros for DMFT compared with a regular Poisson regression model. We believe the lack of goodness-of-fit is due to the heterogeneity in the effects of risk factors on DMFT among population. Thus, our proposed Poisson mixture regression model with subgroup detection can yield better model-fitting results in the present study.

According to our analysis, there was strong statistical evidence that the population can be divided into three groups according to BIC. By comparing the mean DMFT values, we found that the mean DMFT value of subgroup 3 was much higher than that of other two subgroups. With the Poisson mixture regression model which clustered

Table 5 Estimation and 95% CI for Wald test to the effect of sugar intake factor X_{10} under the case of controlling regional factors X_1 and X_2 at different levels in different subgroups

(X_1, X_2)	Subgroup 1			Subgroup 2			Subgroup 3		
	Coef.	SE	CI	Coef.	SE	CI	Coef.	SE	CI
(1 to 0)	0.019	0.001	0.010 to 0.027	0.122	0.011	0.044 to 0.200	0.007	0.001	-0.002 to 0.016
(0 to 1)	0.130	0.005	0.091 to 0.168	0.051	0.001	0.042 to 0.060	0.030	0.001	0.021 to 0.040
(0 to 0)	0.100	0.007	0.050 to 0.149	0.028	0.001	0.018 to 0.038	0.019	0.002	0.008 to 0.031

X_1, X_2 stands for the value of factor X_1 and X_2 , other notations are of the same with [table 2](#).

Table 6 Estimation and 95% CI for Wald test to the effect of the tooth brushing factor X_9 under the case of the utilisation of dental service factors X_{13} and X_{14} at different levels in different subgroups

	Subgroup 1			Subgroup 2			Subgroup 3		
	Coef.	SE	CI	Coef.	SE	CI	Coef.	SE	CI
X_9X_{13}	0.432	0.005	0.422 to 0.442	0.340	0.004	0.332 to 0.348	0.215	0.003	0.209 to 0.221
X_9X_{14}	0.730	0.010	0.710 to 0.750	0.459	0.007	0.445 to 0.473	0.325	0.006	0.313 to 0.337
$X_{13}, X_{14}(1,0)$	-1.174	0.019	-1.311 to 1.036	-0.891	0.013	-0.990 to 0.792	-0.634	0.011	-0.713 to 0.556
$X_{13}, X_{14}(2,0)$	-0.742	0.017	-0.866 to 0.617	-0.551	0.011	-0.633 to 0.469	-0.420	0.009	-0.487 to 0.352
$X_{13}, X_{14}(3,0)$	-0.310	0.017	-0.431 to 0.188	-0.211	0.010	-0.286 to 0.136	-0.205	0.009	-0.271 to 0.139
$X_{13}, X_{14}(4,0)$	0.123	0.017	-0.006 to 0.251	0.129	0.011	0.047 to 0.211	0.010	0.010	-0.062 to 0.082
$X_{13}, X_{14}(5,0)$	0.554	0.020	0.410 to 0.699	0.469	0.013	0.370 to 0.568	0.225	0.012	0.140 to 0.310
$X_{13}, X_{14}(1,1)$	-0.444	0.018	-0.576 to 0.311	-0.433	0.013	-0.531 to 0.334	-0.309	0.011	-0.387 to 0.230
$X_{13}, X_{14}(2,1)$	-0.012	0.016	-0.131 to 0.107	-0.093	0.011	-0.174 to 0.012	-0.094	0.010	-0.161 to 0.027
$X_{13}, X_{14}(3,1)$	0.420	0.016	0.306 to 0.535	0.247	0.010	0.174 to 0.322	0.121	0.009	0.057 to 0.185
$X_{13}, X_{14}(4,1)$	0.853	0.016	0.731 to 0.974	0.587	0.011	0.507 to 0.668	0.335	0.009	0.266 to 0.405
$X_{13}, X_{14}(5,1)$	1.285	0.019	1.148 to 1.422	0.928	0.013	0.830 to 1.025	0.550	0.011	0.468 to 0.633

X_{13} , X_{14} represents the values of X_{13} and X_{14} . 'Coef.' is the estimation of the effect of under different cases, 'SE' stands for the SE of Wald statistics in testing Coef. is 0 or not, and 'CI' is the corresponding 95% CI.

individuals into groups and helped us to find the individualised group, the subgroup 3 was referred to as adolescents with high risk of dental caries. According to the parameter estimates of the three subgroups in the model, we also found that the heterogeneity of risk factors can be explained by social demographic characteristics.

Dental caries is a disease caused by an ecological shift in the composition and activity of the bacterial biofilm when exposed over time to fermentable carbohydrates, leading to a break in the balance between demineralisation and remineralisation.²³ The prevalence of dental caries has been increasing with China's rapid economic development in the past 10 years.⁶ However, due to the distinctions in economic status, culture, education and diet among different regions in China, the prevalence of dental caries differs widely from east to west. This study found that adolescents in the eastern region have fewer DMFT in the average level of the caries population and that adolescents in the central region have fewer DMFT in the high caries population. Adolescents in the western region still had the highest prevalence of dental caries. By the end of 2015,²⁴ there were great regional differences in gross domestic product (GDP). The per capita GDP of the eastern, central and western regions was \$11,400, \$6635 and \$5967, respectively. According to the Report of the National Investigation of Resources for Oral Health in China (2015), the ratio of the number of dentists to the population of the east, centre and west regions was 1:6265, 1:8253 and 1:9968, respectively.²⁵ In addition, the economic development level and the distribution of human resources for oral health in China were extremely uneven. For instance, in

some provinces and municipalities at a higher level of economic development, such as Beijing and Tianjin, the ratio of dentists to the population had already reached the WHO criteria and in some areas (eg, the urban area) had even exceeded the standard of developed countries. However, in underdeveloped areas, this ratio remained much lower than the WHO criteria, and the dental workforce was seriously insufficient. This unequal distribution of economic development levels and human resources for oral health is be a major reason for the regional differences in dental caries.

The rapid growth in regions with high-level economic development has changed lifestyles; more cariogenic foods and drinks are consumed as a result, making sugar the greatest dietary risk factor for the development of dental caries.²⁶ There is a change in behaviour from early childhood to adolescence as adolescents become more independent in selecting their food and drinks. This could increase the risk of caries development.²⁷ As shown in our study, adolescents with a higher frequency of sugar consumption had more DMFT, but the influence decreased with the increasing severity of DMFT. In the case of controlling sugar intake and increasing sugar consumption, the influence of the eastern regional factor gradually weakened, while the difference between the central and western regions became significant and gradually increased. We previously noted that the eastern region had the most highly economically developed areas and greater accessibility of oral health services. This may be why the influence of sugar consumption weakened with increasing sugar consumption in the eastern region. We hypothesised that sugar consumption could have a more

powerful influence in the western or central regions due to regional inequality.

Older adolescents are more likely to have more DMFT. Clinical study data show that newly erupted teeth, especially the first permanent molars, are most vulnerable to dental caries in the first 2–4 years.²⁸ Our study showed that the influence of age was strongest in the high caries subgroup. Therefore, the importance of focusing on preventive measures in the first critical years after tooth eruption in the high caries risk population based on the caries risk assessment system should be emphasised.²⁹

As found in previous studies,^{30–31} pit-and-fissure sealant could prevent caries. Approximately 60.9% of dental caries in the permanent teeth of 12-year-old Chinese children occurs in the pits and fissures of posterior teeth.² Despite strong evidence of effectiveness, sealants were underused, especially among adolescents at higher risk for dental caries.³² Less than 7% of 12-year-old Chinese adolescents have received sealants,² which is far below the level of developed countries.³³ US national data indicate that 38% of this population received sealants, compared with 47% of higher-income children. Increasing the prevalence of sealants among children is a national health objective in China.³⁴ Moreover, our study found that the power of sealants for preventing DMFT decreased with increasing severity. This is an indication that sealant is not sufficient for the prevention of caries in high-risk adolescents. These adolescents must be managed aggressively to eliminate or reduce the possibility of new or recurrent caries lesions. For example, bacterial testing, antimicrobial treatments, 1.1% NaF toothpaste, 5% NaF fluoride varnish and xylitol are standard regimens for all high-risk patients.³⁵

This study has some limitations. This research on dental caries using cross-sectional data collection could not represent casual relationships between dental caries and the related factors, and hence, was unable to indicate how caries changes over time in the general population or in a given sample and what variables are associated with the overall trend and the differences within and among persons. In addition, dental caries data with larger scale and longitudinal comparisons would be necessarily beneficial to validate the model-fitting results by the Poisson mixture regression model in future studies. Nevertheless, the latest data used in the present study could provide some helpful suggestions for the future trend of oral health promotion measures in China. Owing to the heterogeneity of risk factors in different populations, group-individualised strategies based on risk assessment levels are needed for adolescents with distinct dental health characteristics and will affect decisions in the use of antibiotics, fluoride, sealants, and the frequency of radiographs and periodic oral examinations, as well as management procedures for other risk factors. Public health policy makers should also keep in mind that group individualisation is of important value to be carefully considered in planning future population-based oral health promotion programmes.

In summary, the risk factors of dental caries exhibited heterogeneity in groups with different characteristics. The Poisson mixture regression model could cluster individuals into groups and identify the heterogeneous effects of risk factors among different groups. The findings support the need for different targeted interventions and prevention measures in groups with different dental health characteristics.

Acknowledgements We are very grateful to all the participants, examiners and interviewers in the Fourth National Oral Health Survey.

Contributors CY, JH, JK, SZ contributed to conception, design, data acquisition, analysis, and interpretation, drafted and critically revised the manuscript; XS contributed to conception, design, and interpretation, drafted and critically revised the manuscript; all authors gave final approval and agree to be accountable for all aspects of the work.

Funding This study was supported by 'Scientific Research Fund of National Health Commission of the People's Republic of China (201502002)'.

Competing interests None declared.

Patient consent for publication Not required.

Ethics approval Ethical clearance was approved by the Stomatological Ethics Committee of the Chinese Stomatological Association (approval number 2014-003).

Provenance and peer review Not commissioned; externally peer reviewed.

Data availability statement Data are available upon reasonable request. Data are available upon reasonable request. E-mail address: kqzsg86@bjmu.edu.cn.

Supplemental material This content has been supplied by the author(s). It has not been vetted by BMJ Publishing Group Limited (BMJ) and may not have been peer-reviewed. Any opinions or recommendations discussed are solely those of the author(s) and are not endorsed by BMJ. BMJ disclaims all liability and responsibility arising from any reliance placed on the content. Where the content includes any translated material, BMJ does not warrant the accuracy and reliability of the translations (including but not limited to local regulations, clinical guidelines, terminology, drug names and drug dosages), and is not responsible for any error and/or omissions arising from translation and adaptation or otherwise.

Open access This is an open access article distributed in accordance with the Creative Commons Attribution Non Commercial (CC BY-NC 4.0) license, which permits others to distribute, remix, adapt, build upon this work non-commercially, and license their derivative works on different terms, provided the original work is properly cited, appropriate credit is given, any changes made indicated, and the use is non-commercial. See: <http://creativecommons.org/licenses/by-nc/4.0/>.

ORCID iD

Shuguo Zheng <http://orcid.org/0000-0002-4156-1619>

REFERENCES

- Chulani VL, Gordon LP. Adolescent growth and development. *Prim Care* 2014;41:465–87.
- Yuan C, Wang XZ, Sun XY, *et al*. Oral health status of 12-year-olds from regions with and without coverage of the National oral health comprehensive intervention program for children in China. *Chin J Dent Res* 2018;21:299–306.
- GBD 2016 Disease and Injury Incidence and Prevalence Collaborators. Global, regional, and national incidence, prevalence, and years lived with disability for 328 diseases and injuries for 195 countries, 1990–2016: a systematic analysis for the global burden of disease study 2016. *Lancet* 2017;390:1211–59.
- Qi X. *Report of the third national oral health survey of China*. Beijing: People's Medical Publishing House, 2008.
- Hu D-Y, Hong X, Li X. Oral health in China-trends and challenges. *Int J Oral Sci* 2011;3:7–12.
- Quan JK, Wang XZ, Sun XY, *et al*. Permanent teeth caries status of 12- to 15-year-olds in China: findings from the 4th national oral health survey. *Chin J Dent Res* 2018;21:181–93.
- Shen A, Zeng X, Cheng M, *et al*. Inequalities in dental caries among 12-year-old Chinese children. *J Public Health Dent* 2015;75:210–7.
- Peng B, Petersen PE, Fan MW, *et al*. Oral health status and oral health behaviour of 12-year-old urban schoolchildren in the people's Republic of China. *Community Dent Health* 1997;14:238–44.



- 9 Zhu L, Petersen PE, Wang H-Y, *et al.* Oral health knowledge, attitudes and behaviour of children and adolescents in China. *Int Dent J* 2003;53:289–98.
- 10 Wolfe JH. A computer program for the maximum likelihood analysis of types. naval personnel research activity San Diego Ca 1965.
- 11 Banfield JD, Raftery AE. Model-based Gaussian and non-Gaussian clustering. *Biometrics* 1993;49:803–21.
- 12 McLachlan GJ, Basford KE. *Mixture models: inference and applications to clustering*: M. Dekker New York, 1988.
- 13 Fraley C, Raftery AE. Model-Based clustering, discriminant analysis, and density estimation. *J Am Stat Assoc* 2002;97:611–31.
- 14 Manthey J, Shield KD, Rylett M, *et al.* Global alcohol exposure between 1990 and 2017 and forecasts until 2030: a modelling study. *Lancet* 2019;393:2493–502.
- 15 Kassebaum NJ, Bernabé E, Dahiya M, *et al.* Global burden of untreated caries: a systematic review and metaregression. *J Dent Res* 2015;94:650–8.
- 16 Xu M, Cheng M, Gao X, *et al.* Factors associated with oral health service utilization among adults and older adults in China, 2015–2016. *Community Dent Oral Epidemiol* 2020;48:32–41.
- 17 China NBoSo. Tabulation of the 2010 population census of the people's republic of China, 2010. Available: <http://www.stats.gov.cn/tjsj/pcsj/rkpc/6rp/indexch.htm>
- 18 Grün B, Leisch F. Applications of finite mixtures of regression models, 2007. Available: <http://cran.r-project.org/web/packages/flexmix/vignettes/regression-examples.pdf>
- 19 Leisch F. *Flexmix: a general framework for finite mixture models and latent class regression in R*, 2004.
- 20 Grün B, Leisch F. *FlexMix version 2: finite mixtures with concomitant variables and varying and constant parameters*, 2008.
- 21 Chamberlain SR, Stochl J, Redden SA, *et al.* Latent class analysis of gambling subtypes and impulsive/compulsive associations: time to rethink diagnostic boundaries for gambling disorder? *Addict Behav* 2017;72:79–85.
- 22 Tran G, Hensor EMA, Ray A, *et al.* Ultrasound-detected pathologies cluster into groups with different clinical outcomes: data from 3000 community referrals for shoulder pain. *Arthritis Res Ther* 2017;19:30.
- 23 Young DA, Nový BB, Zeller GG, *et al.* The American dental association caries classification system for clinical practice: a report of the American dental association Council on scientific Affairs. *J Am Dent Assoc* 2015;146:79–86.
- 24 NHC. *National health statistical yearbook 2015*. Beijing: National Health Commission of China, 2016.
- 25 Sun XY, Yuan C, Wang XZ, *et al.* Report of the National investigation of resources for oral health in China. *Chin J Dent Res* 2018;21:285–97.
- 26 Sheiham A. Dietary effects on dental diseases. *Public Health Nutr* 2001;4:569–91.
- 27 Peres MA, Sheiham A, Liu P, *et al.* Sugar consumption and changes in dental caries from childhood to adolescence. *J Dent Res* 2016;95:388–94.
- 28 Lynch RJM. The primary and mixed dentition, post-eruptive enamel maturation and dental caries: a review. *Int Dent J* 2013;63 Suppl 2:3–13.
- 29 Rechmann P, Chaffee BW, Rechmann BMT, *et al.* Caries management by risk assessment: results from a practice-based research network study. *J Calif Dent Assoc* 2019;47:15–24.
- 30 Ahovuo-Saloranta A, Forss H, Walsh T, *et al.* Sealants for preventing dental decay in the permanent teeth. *Cochrane Database Syst Rev* 2013;Cd001830.
- 31 Ahovuo-Saloranta A, Forss H, Walsh T, *et al.* Pit and fissure sealants for preventing dental decay in permanent teeth. *Cochrane Database Syst Rev* 2017;7:Cd001830.
- 32 Beauchamp J, Caufield PW, Crall JJ, *et al.* Evidence-Based clinical recommendations for the use of pit-and-fissure sealants: a report of the American dental association Council on scientific Affairs. *J Am Dent Assoc* 2008;139:257–68.
- 33 Griffin S, Naavaal S, Scherrer C, *et al.* School-Based dental sealant programs prevent cavities and are cost-effective. *Health Aff* 2016;35:2233–40.
- 34 Wang C, Rao K, Wu S, *et al.* Health care in China: improvement, challenges, and reform. *Chest* 2013;143:524–31.
- 35 Weintraub JA, Ramos-Gomez F, Jue B, *et al.* Fluoride varnish efficacy in preventing early childhood caries. *J Dent Res* 2006;85:172–6.