One Health 19 (2024) 100889

Contents lists available at ScienceDirect

One Health

journal homepage: www.elsevier.com/locate/onehlt

Application of compound poisson model to estimate underreported risk of non-communicable diseases in underdeveloped areas

Hongli Wan [a,1], Wenhui Zhu [a,1], Jingmin Yan [a], Xinyue Han [a], Jie Yu [a], Qiang Liao [b], Tao Zhang [a,*]

[a] Department of Epidemiology and Health Statistics, West China School of Public Health and West China Fourth Hospital, Sichuan University, Chengdu 610041, Sichuan Province, China

[b] Liangshan Prefecture Center for Disease Control and Prevention, Xichang 615000, Sichuan Province, China

A R T I C L E I N F O

Keywords:
Bayesian methods
Compound Poisson model
Underreporting
Hypertension
Type 2 diabetes

A B S T R A C T

Background: Hypertension and diabetes are major components of non-communicable diseases (NCDs), with a substantial number of patients residing in underdeveloped areas. Limited medical resources in these areas often results in underreporting of disease prevalence, masking the true extent of diseases. Taking the underdeveloped Liangshan Yi Autonomous Prefecture in China as an example, this study aimed to correct the underreported prevalence of hypertension and type 2 diabetes so as to provide inspiration for the allocation of medical resources in such areas.

Methods: Assuming the true number of patients in each area follows a Poisson distribution, we applied a Compound Poisson Model based on Clustering of Data Quality (CPM-CDQ) to estimate the potential true prevalence of hypertension and diabetes, as well as the registration rate of existing patients. Specifically, a hierarchical clustering approach was utilized to group the counties based on the data quality, and then the registration rate of the cluster with the best data quality was used as a priori information for the model. The model parameters were estimated by the maximum likelihood method. Sensitivity analyses were performed to test the robustness of the model.

Results: The estimated prevalence of hypertension in the entire Liangshan Prefecture from 2018 to 2020 ranged from 24.59 % to 25.28 %, and for diabetes, it ranged from 4.95 % to 8.42 %. The registration rates for hypertension and diabetes were 14.10 % to 24.59 % and 15.98 % to 29.12 %, respectively. Additionally, the accuracy of clustering the counties with the best data quality had a significant impact on the performance of the model.

Conclusion: Liangshan Prefecture is experiencing a significant high prevalence of hypertension and diabetes, accompanied by a concerningly low registration rate. The CPM-CDQ proved useful for assessing underreporting risks and facilitating targeted interventions for NCDs control and prevention, particularly in underdeveloped areas.

1. Introduction

Non-communicable diseases (NCDs) are responsible for an estimated 41 million deaths annually, accounting for 71 % of all global deaths [1,2]. This issue is more severe in underdeveloped areas, where seven out of every ten deaths in 2020 were attributed to NCDs [1,2]. Hypertension and diabetes represent two significant global public health problems, with approximately 1.28 billion adults aged 30–79 suffering from hypertension and 529 million people living with diabetes. Notably, two-thirds of hypertensive patients and four-fifths of type 2 diabetes

Abbreviations: CPM-CDQ, Compound Poisson Model based on Clustering of Data Quality;; CPM, Compound Poisson model;; NCDs, Non-communicable diseases;; Liangshan Prefecture, Liangshan Yi Autonomous Prefecture;; CSS, Cross-sectional study;; CMR, Capture-mark-recapture;; CCDRFSS, China Chronic Disease and Risk Factors Surveillance System;; CDC, Center for Disease Control and Prevention;; log (MSE), logarithmic mean square error;; 95 % *PI*, 95 % posterior prediction intervals.

* Corresponding author at: Department of Epidemiology and Health Statistics, West China School of Public Health and West China Fourth Hospital, Sichuan University, No. 16, Section 3, Renmin South Road, Chengdu, Sichuan CN 610041, China.

E-mail address: statzhangtao@scu.edu.cn (T. Zhang).

[1] These authors contributed equally to this work as first authors.

https://doi.org/10.1016/j.onehlt.2024.100889
Received 13 December 2023; Received in revised form 8 August 2024; Accepted 5 September 2024
Available online 8 September 2024

patients reside in underdeveloped areas [3–6]. Given the long disease cycles and complex etiologies of NCDs, successful prevention and treatment heavily depend on individualized and refined health management services [7]. The Chronic Care Model proposed by the American Diabetes Association, encompassing six systemic components [8,9], emphasized personalized and refined management of NCDs in developed countries [10]. However, in less developed countries, such comprehensive systems and advanced technologies of management are often problematic. In low- and middle-income countries, fewer than one-third of hypertensive patients receive antihypertensive medication [11,12]. Given the significant societal resources required to control and prevent NCDs, underdeveloped areas face a monumental challenge in resource allocation.

For better allocation of resources, it is crucial to understand the distribution of prevalence rates, which are typically obtained through disease surveillance systems. However, the underreporting poses a significant challenge in estimating the true prevalence of NCDs. For instance, an Australian health survey found underreporting of approximately 16 % for hypertension and 1.3 % for diabetes [13]. While research on underreporting in underdeveloped areas remains scarce, existing evidence hints at a potentially larger problem. Factors contributing to underreporting include inconsistent data formats due to a lack of standardized data collection guidelines [14,15], legal restrictions, limited health facilities usage [16], insufficient healthcare personnel [17], and low health literacy levels [18]. These problems are exacerbated in medically underprivileged areas with limited healthcare resources. Therefore, accurately assessing the potential true number of patients is crucial for underdeveloped areas to implement targeted intervention measures.

For methods of estimating the prevalence of NCDs, commonly used methods include the large-sample cross-sectional survey(CSS) [14 [16–18] and the capture-mark-recapture (CMR). Large-sample epidemiological surveys often require large human and material resources, and are difficult to organize and implement. The CMR requires a relatively stable population during the study period [19,20], which is impractical in underdeveloped areas with high population mobility.

When the complete number of patients is not available, the incomplete counts that have been observed could represent the minimum of the true number of patients and provide some information for the estimation of the true prevalence. By assuming that the potential true number of patients followed a Poisson distribution [21] and the number of registered patients followed a binomial distribution, Stoner et al. [22] proposed a method for estimating the true prevalence rate based on the observed incomplete number of patients, namely the compound Poisson model (CPM), which was demonstrated as a robust framework for analyzing potentially underreported count data via the application case of Brazilian tuberculosis. Here, "robust" refers to the model's capacity to avoid significant systematic bias under different data conditions. The model incorporates the average underreporting rate across the research areas as prior to supplement the missing data and considers random effects in the disease occurrence and underreporting processes. However, obtaining the average underreporting rate is usually difficult in underdeveloped areas.

In such a circumstance, the Compound Poisson Model based on Clustering of Data Quality (CPM-CDQ) [23] is a more preferable choice, because it shifts the CPM's reliance on the average data integrity rate across the research areas to a requirement for the data integrity rate of only areas with the best data quality. In order to map the underreporting risk of early neonatal mortality (ENM) in Minas Gerais State (MG), Brazil, Oliveira et al. [23] utilized information on the reporting rates in areas with the best data quality as a priori for the model, demonstrating an effective performance of CPM-CDQ. This provided a fantastic and enlightening idea for estimating the prevalence of hypertension and diabetes in underdeveloped areas.

Following the inspiring work of Oliveira et al. [23], this study aimed to offer a template on how existing methodologies for underreporting estimation could be applied in other underdeveloped areas. We took Liangshan Yi Autonomous Prefecture (Liangshan Prefecture) in Sichuan Province of China as an application example. Liangshan Prefecture is an underdeveloped, multi-ethnic, and mountainous area. Previous studies have confirmed that the registration of NCDs patients in Liangshan Prefecture is incomplete due to misdiagnosis or underreporting [24–26]. By analyzing the underreporting risk in Liangshan Prefecture NCDs data, some clues could be provided for estimating the risk of underreporting, identifying the local priority prevention and control areas, and implementing targeted measures. We also conducted sensitivity analyses to assess the applicability of CPM-CDQ under different data conditions.

2. Methods

In this section, we described a step-by-step implementation of the CPM-CDQ proposed by Oliveira et al. [23], encompassing both the model and sensitivity analyses. The analysis process of the CPM-CDQ was depicted in Fig. 1. For a comprehensive understanding of the CPM-CDQ, we also provided detailed descriptions of the CPM and CPM-CDQ in Supplementary Materials 1–2.

2.1. Compound poisson model based on clustering of data quality

Step 1: Data preprocessing.

The counting data of hypertension and type 2 diabetes patients in 17 counties of Liangshan Prefecture, Sichuan Province, from 2018 to 2020 were extracted from the China Chronic Disease and Risk Factors Surveillance System (CCDRFSS). The registration process involved: (1) diagnosis in medical institutions followed by information forwarding to township health centers for registration; (2) information was registered in CCDRFSS by the CDC, with subsequent follow-up management by health centers.

The covariates were sourced from the Statistical Yearbook of Liangshan Prefecture (http://tjj.sc.gov.cn/) and were centered (i.e., each variable's value was subtracted by its mean). For details on the specific variables used, please refer to Supplementary Material 3. Supplementary Material 4 presented the number of permanent residents aged 15 years and above in each county in Liangshan Prefecture from 2018 to 2020, along with the distribution of factors related to disease levels and registration rates.

Step 2: Priors based on regional clustering.

Based on the covariates, divisive hierarchical clustering was performed on the counties. This method created a hierarchical structure that automatically split clusters without pre-specifying the number. Subsequently, the clusters were ranked based on the data reporting quality, assessed by the completeness or registration rate of patient information. Furthermore, the clustering of counties was adjusted based on the expertise of local public health experts involved in patient registration. The mean and variance of the reporting rate of cluster with the best data quality were adopted as the prior distribution (known as the conditional uniform distributions). Detailed information was provided in the Supplementary Material 5.

Step 3: Estimating potential patients.

The prior distribution and the observed number of patients of each county were input into the CPM-CDQ. The posterior distribution was estimated using the Markov Chain Monte Carlo. Setting the burn-in to 50,000 iterations, the model was run with 100,000 samples until convergence was achieved. The output of the model was the annual estimate of potential patients for each county.

Step 4: Estimating the prevalence and registration rate.

The potential prevalence rate was calculated by dividing the estimated number of potential patients by the total population. The underreporting risk was assessed by the ratio of the difference between potential patients and observed cases to the total number of potential patients.

Step 5: Sensitivity analyses.

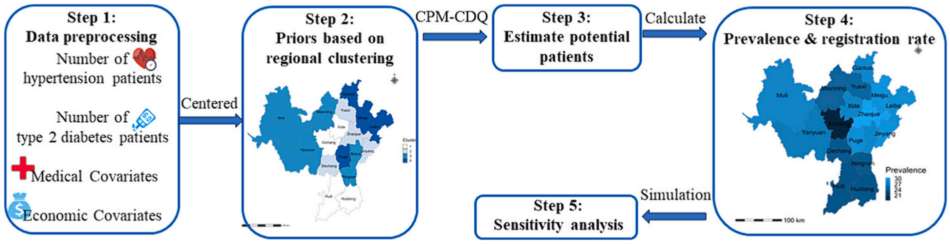

Fig. 1. Modeling workflow.

In order to investigate the robustness of the CPM-CDQ in assessing underreporting risk under varying data conditions, following Oliveira et al. [23], sensitivity analyses were performed with simulated data. Factors such as the accuracy of prior information, data completeness level, regional clustering accuracy, and cluster size were varied. The evaluation metrics included prediction bias, logarithmic mean square error (log(MSE)), and the coverage of the 95 % posterior prediction intervals (95 % *PI*) against actual values. Specific details were available in Supplementary Material 6.

2.2. Application example

The observed patient numbers were obtained directly from the CDC without model adjustment. The potential number of patients was estimated using the CPM-CDQ, based on the observed numbers and covariates. Registration and underreporting rates were calculated based on these estimated potential patient numbers.

3. Results

3.1. Descriptive statistics

The Figs. 2 and 3 presented the spatial distribution maps of hypertension and diabetes prevalence in Liangshan prefecture from 2018 to 2020, with county-level prevalence rates ranging from approximately 2.57 % to 8.92 % and 0.88 % to 1.82 % , respectively. However, in reality, the prevalence of hypertension in this area was approximately 25 %, according to the results of previous field surveys [27]. This discrepancy indicated a potential underestimation in the number of actual patients. The predicted prevalence rates exhibited distinct spatial and temporal characteristics. Spatially, the northeastern region of Liangshan Prefecture, including Zhaojue, Meigu, Puge, Butuo, Xide, Mianning, Yuexi, Ganluo, and Jinyang counties, demonstrated relatively higher prevalence levels.

3.2. Estimation of the potential prevalence and registration rate of hypertension and diabetes

3.2.1. Covariates

Age and urbanization rate were recognized as influential factors for hypertension and diabetes. Consequently, data on the aging rate, urbanization rate, and standardized management rate of the elderly population for each county were collected [27–29]. Additionally, per capita medical consumption expenditure and the medical personnel density were gathered, given that the registration rates in Liangshan Prefecture from 2018 to 2020 were significantly adjusted due to data integration and the impact of the COVID-19 pandemic, which directly influenced the reporting of hypertension and diabetes cases [30,31].

3.2.2. Priors based on regional clustering

The cluster analysis was conducted based on these variables to assess the registration rates. To mitigate the impact of variable scales on clustering outcomes, data standardization was performed prior to analysis. Hierarchical clustering categorized the 17 counties into four distinct clusters, as depicted in Fig. 4. The Cluster 1, representing the highest data quality, was characterized by the lowest underreporting rate. The average underreporting rate of Cluster 1 was utilized as the prior of the CPM-CDQ model (detailed information was provided in Supplementary Material 5).

3.2.3. Model checking

Figs. SA1 to SA3 (Supplementary Material 7) displayed the distribution of the number of hypertension and diabetes patients per 1000 people, alongside the corresponding number of patients fitted by the CPM-CDQ. The scattered points predominantly aligned along a straight line with a slope of 1, signifying a robust model fit to the registration data. The 95 %*CI* of the difference between the CPM-CDQ estimated registered patients and the observed patient count was symmetrically distributed around zero, indicating no systematic deviation of the model.

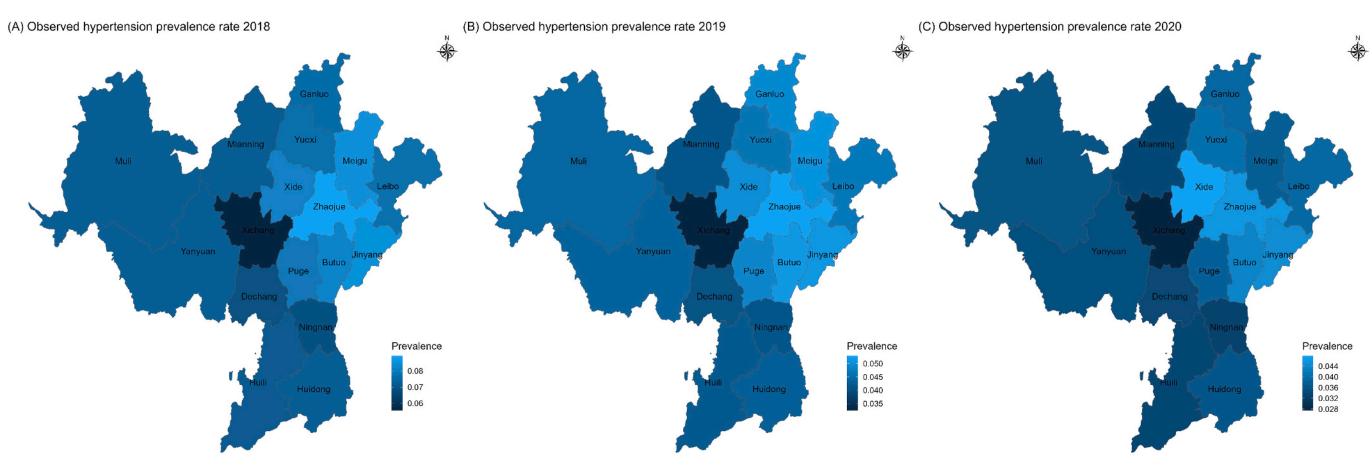

Fig. 2. Prevalence rate of hypertension from 2018 to 2020 (The base map was from http://bzdt.ch.mnr.gov.cn/).

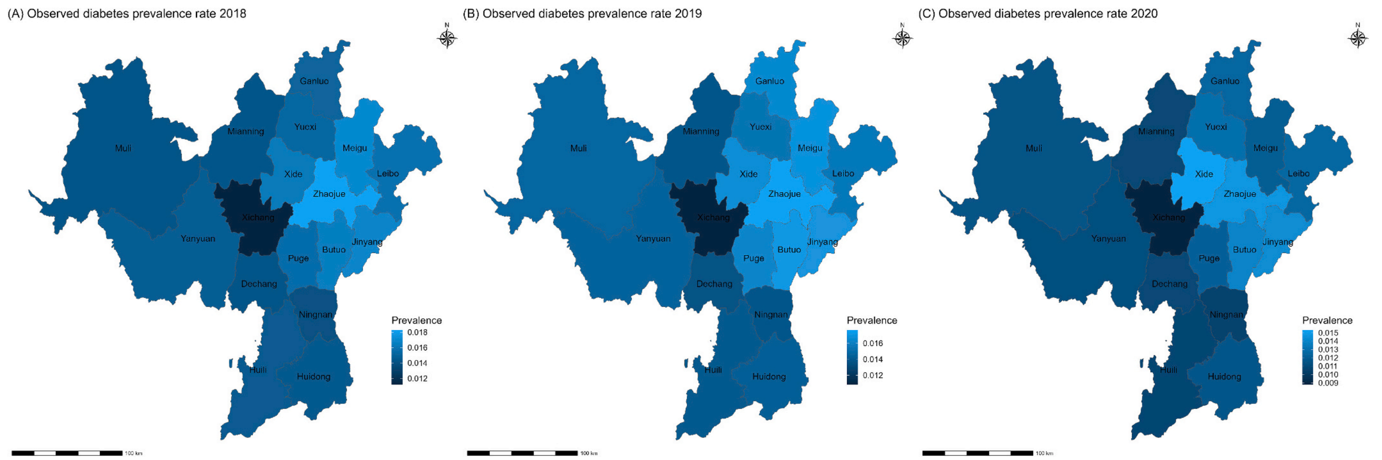

Fig. 3. Prevalence rate of diabetes from 2018 to 2020 (The base map was from http://bzdt.ch.mnr.gov.cn/).

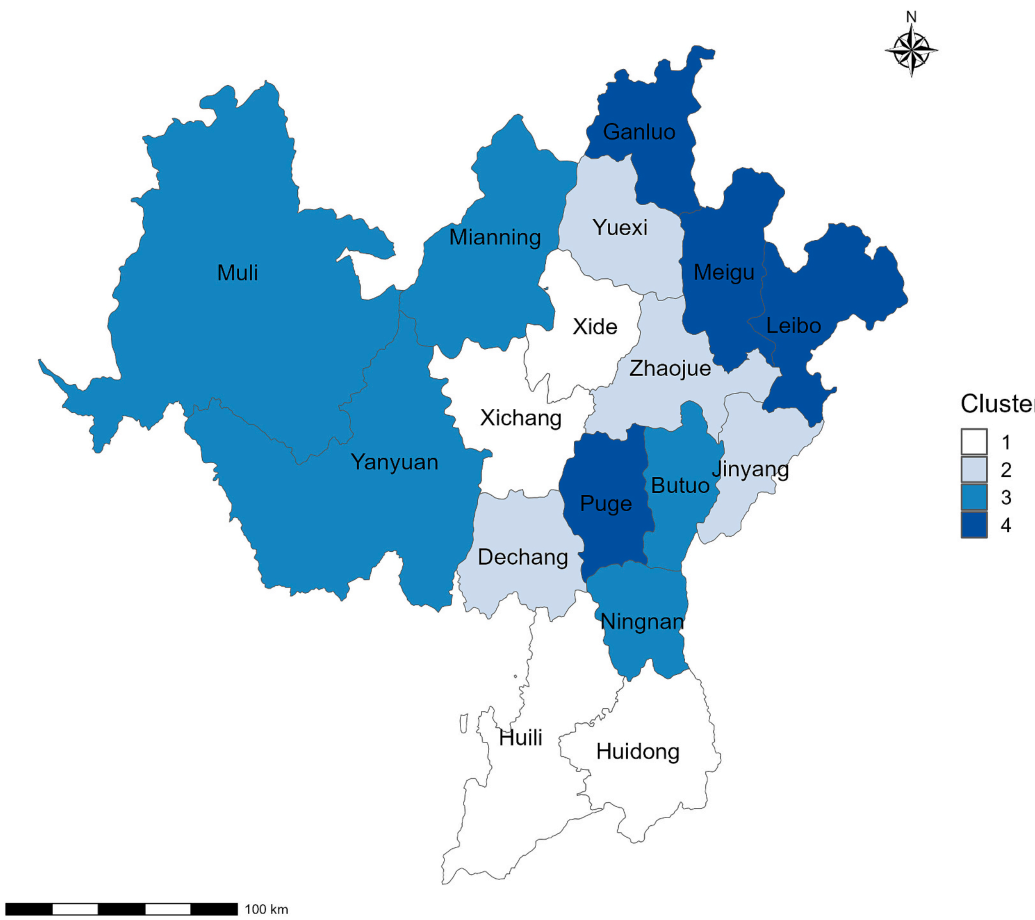

Fig. 4. Clustered results (The used base map was from http://bzdt.ch.mnr.gov.cn/).

Figs. SA4 to SA9 illustrated the prior and posterior distributions of sample statistics such as mean, variance, and mean squared error. The posterior predictive distributions exhibited reduced variability compared to the prior, indicating that the model effectively captured the posterior information of the observed values. The posterior distribution of mean squared error showed a narrowed variability range, suggesting enhanced predictive accuracy.

3.2.4. *Potential prevalence and registration rates*

Tables 1 and 2 presented the estimated and observed hypertension and diabetes prevalence rates and registration rates for each county from 2018 to 2020. The estimated hypertension prevalence rates in the entire Liangshan Prefecture were 24.59 % (95 % *CI*: 23.73 %, 25.98 %), 25.21 % (95 % *CI*: 24.42 %, 26.00 %), and 25.28 % (95 % *CI*: 23.31 %, 26.15 %) for the respective years. The estimated diabetes prevalence rates in the entire Liangshan Prefecture were 4.95 % (95 % *CI*: 4.80 %, 5.11 %), 8.42 % (95 % *CI*: 8.10 %, 8.75 %), and 7.12 % (95 % *CI*: 6.82 %, 7.42 %) for the respective years. Fig. 5 illustrated the overall trends in potential prevalence and registration rates for hypertension and diabetes in Liangshan Prefecture, showing no significant overall increase or

Table 1
Estimated hypertension prevalence and registration rate of counties.

County	2018			2019			2020		
	Prevalence[a] (%)	CPM-CDQ		Prevalence[a] (%)	CPM-CDQ		Prevalence[a] (%)	CPM-CDQ	
		Prevalence[b] (5th and 95th percentiles)(%)	Registration rate (5th and 95th percentiles) (%)		Prevalence[b] (5th and 95th percentiles)(%)	Registration rate (5th and 95th percentiles) (%)		Prevalence[b] (5th and 95th percentiles)(%)	Registration rate (5th and 95th percentiles) (%)
Xichang	5.57	18.27(17.96, 18.56)	30.48(30.05, 30.95)	3.24	18.53(18.04, 19.07)	17.52(17.06, 17.96)	2.75	18.95(18.34, 19.58)	14.51(14.06, 14.95)
Muli	7.25	27.94(26.12, 27.79)	26.90(26.22, 27.63)	4.41	26.45(25.53, 27.43)	16.66(16.20, 17.09)	3.61	26.86(25.75, 28.02)	13.46(13.02, 13.87)
Yanyuan	7.28	27.05(26.30, 27.85)	26.90(26.22, 27.63)	4.36	26.17(25.39, 27.02)	16.66(16.20, 17.09)	3.51	26.08(25.16, 27.05)	13.46(13.02, 13.87)
Dechang	6.93	24.14(23.53, 24.76)	28.74(28.14, 29.40)	4.10	23.98(23.24, 24.79)	17.09(16.63, 17.51)	3.43	24.86(23.95, 25.82)	13.80 (13.36, 14.23)
Huili	7.18	23.57(23.16, 23.97)	30.48(30.05, 30.95)	4.22	24.07(23.39, 24.81)	17.52(17.06, 17.96)	3.36	23.16(22.38, 23.99)	14.51(14.06, 14.95)
Huidong	7.24	23.77(23.33, 24.19)	30.48(30.05, 30.95)	4.30	24.55(23.84, 25.32)	17.52(17.06, 17.96)	3.66	25.20 (24.35, 26.1)	14.51(14.06, 14.95)
Ningnan	6.92	24.10(23.49, 24.74)	28.74(28.14, 29.40)	4.17	25.01(24.19, 25.87)	16.66(16.20, 17.09)	3.32	24.06(23.15, 25.03)	13.80 (13.36, 14.23)
Puge	7.93	28.90(26.44, 30.55)	26.98(24.05, 28.95)	4.89	30.08(29.10, 31.15)	16.25(15.79, 16.68)	3.83	26.37(25.34, 27.41)	14.51(14.06, 14.95)
Butuo	8.24	28.24(27.53, 29.00)	29.18(28.58, 29.83)	5.15	30.88(29.85, 31.95)	16.66(16.20, 17.09)	4.37	30.83(29.67, 32.05)	14.15(13.71, 14.58)
Jinyang	8.53	29.22(28.49, 29.98)	29.18(28.58, 29.83)	5.12	29.94(28.98, 30.99)	17.09(16.63, 17.51)	4.48	31.62(30.41, 32.87)	14.15(13.71, 14.58)
Zhaojue	8.92	30.57(29.84, 31.35)	29.18(28.58, 29.83)	5.29	30.95(30.01, 31.98)	17.09(16.63, 17.51)	4.62	32.59(31.41, 33.8)	14.15(13.71, 14.58)
Xide	8.19	28.08(27.37, 28.82)	29.18(28.58, 29.83)	5.01	28.57(27.66, 29.56)	17.52(17.06, 17.96)	4.79	33.77(32.51, 35.08)	14.15(13.71, 14.58)
Mianning	7.21	25.08(24.50, 25.70)	28.74(28.14, 29.40)	4.16	25.00(24.25, 25.81)	16.66(16.20, 17.09)	3.42	24.78(23.92, 25.69)	13.80 (13.36, 14.23)
Yuexi	7.72	25.32(24.83, 25.80)	30.48(30.05, 30.95)	4.64	27.15(26.33, 28.02)	17.09(16.63, 17.51)	4.04	27.86(26.9, 28.87)	14.51(14.06, 14.95)
Ganluo	7.62	28.33(27.51, 29.20)	26.90(26.22, 27.63)	4.96	30.53(29.53, 31.57)	16.25(15.79, 16.68)	3.91	29.03(27.94, 30.17)	13.46(13.02, 13.87)
Meigu	8.50	29.14(28.43, 29.89)	29.18(28.58, 29.83)	5.06	31.10(30.12, 32.15)	16.25(15.79, 16.68)	3.78	26.71(25.71, 27.77)	14.15(13.71, 14.58)
Leibo	7.74	28.77(27.94, 29.62)	26.90(26.22, 27.63)	4.70	28.91(27.97, 29.89)	16.25(15.79, 16.68)	3.94	29.31(28.23, 30.41)	13.46(13.02, 13.87)
Liangshan	7.17	24.59(23.73, 25.98)	29.15(27.29, 31.61)	3.10	25.21(24.42, 26.00)	16.98(16.46, 17.52)	2.57	25.28(24.31, 26.15)	14.10(13.58, 14.61)

[a] Observed prevalence rate.
[b] Estimated prevalence rate.

decrease in hypertension prevalence from 2018 to 2020, while registration rates for both diseases exhibited a declining trend. The regional distribution of prevalence and registration rates was detailed in Figs. SA10 and SA11, revealing higher hypertension prevalence in certain northeastern counties, such as Zhaojue, Meigu, Butuo, Jinyang, Puge, and Leibo, with rates ranging from approximately 30 % to 33 %. In contrast, other counties like Xichang, Dechang, Huili, Huidong, Ningnan, and Mianning exhibited lower rates, ranging from 18 % to 24 %. A similar geographic pattern was observed in the prevalence of diabetes.

During the study period, the overall registration rate for hypertension in Liangshan Prefecture exhibited a significant decline. In 2018, the rate was 29.15 % (95 % *CI*: 27.29 %, 31.61 %), which decreased to 16.98 % (95 % *CI*: 16.46 %, 17.52 %) in 2019, and further to 14.10 % (95 % *CI*: 13.58 %, 14.61 %) in 2020. The registration rate for diabetes followed a parallel downward trend, with rates of 29.12 % (95 % CI: 28.20 %, 30.07 %) in 2018, 16.94 % (95 % CI: 16.30 %, 17.61 %) in 2019, and 15.98 % (95 % CI: 15.32 %, 16.67 %) in 2020. The spatial distribution of hypertension registration rates revealed notable disparities across Liangshan Prefecture. In the northeastern and northwestern counties, Puge, Muli, Ganluo, Meigu, and Leibo reported relatively lower rates, ranging from approximately 26 % to 28 %, 15 % to 16 %, and 12 % to 13 %, respectively. Conversely, in Xichang, Huili, Huidong, and Dechang counties, the rates were comparatively higher, with values ranging from approximately 29 % to 31 %, 16 % to 18 %, and 14 % to 15 % over the three-year span. The diabetes registration rates reflected this regional pattern, with lower rates observed in the northeastern area.

3.3. Study on the robustness of the CPM-CDQ

3.3.1. Different prior information accuracy

The model's sensitivity to the accuracy of prior information was assessed by varying the expectation of γ_1 (i.e., the best data quality underreporting rate). Fig. 6 presented the model's fitting error and the 95 % *PI* under different values of γ_1. With a simulated reporting rate of 30 %, the true value of γ_1 was 0.7. An increase in the expected deviation of the prior distribution led to a higher fitting error and a lower 95 % *PI* for the actual values, underscoring the model's sensitivity to prior accuracy.

3.3.2. Different levels of data incompleteness

Fig. 7 illustrated the prediction error of the model under different registration rates and prior mean values. As the registration rate increased, the prediction error logarithmic mean squared error (log (MSE)) of the model decreased gradually under the prior mean values close to the actual. When the registration rate remained the same, a larger deviation of the prior mean from the actual value resulted in a higher log (MSE) produced by the model.

Table 2
Estimated diabetes prevalence and registration rate of counties.

County	2018			2019			2020		
	Prevalence[a] (%)	CPM-CDQ		Prevalence[a] (%)	CPM-CDQ		Prevalence[a] (%)	CPM-CDQ	
		Prevalence[b] (5th and 95th percentiles)(%)	Registration rate (5th and 95th percentiles) (%)		Prevalence[b] (5th and 95th percentiles)(%)	Registration rate (5th and 95th percentiles) (%)		Prevalence[b] (5th and 95th percentiles)(%)	Registration rate (5th and 95th percentiles) (%)
Xichang	1.12	3.67(3.59, 3.76)	30.50(30.05, 30.95)	1.08	6.20(6.00, 6.40)	17.49(17.05, 17.95)	0.88	5.34(5.16, 5.53)	16.49(16.04, 16.95)
Muli	1.43	4.98(4.77, 5.19)	28.75(28.11, 29.36)	1.47	8.84(8.43, 9.26)	16.63(16.19, 17.08)	1.16	7.38(6.99, 7.79)	15.68(15.24, 16.12)
Yanyuan	1.47	5.12(4.96, 5.28)	28.75(28.11, 29.36)	1.45	8.74(8.42, 9.07)	16.63(16.19, 17.08)	1.12	7.16(6.88, 7.46)	15.68(15.24, 16.12)
Dechang	1.42	4.88(4.71, 5.05)	29.18(28.54, 29.81)	1.37	8.02(7.69, 8.36)	17.06(16.63, 17.51)	1.1	6.84(6.53, 7.15)	16.08(15.65, 16.53)
Huili	1.46	4.79(4.67, 4.91)	30.50(30.05, 30.95)	1.41	8.04(7.76, 8.33)	17.49(17.05, 17.95)	1.08	6.53(6.28, 6.78)	16.49(16.04, 16.95)
Huidong	1.45	4.76(4.64, 4.89)	30.50(30.05, 30.95)	1.43	8.20(7.91, 8.50)	17.49(17.05, 17.95)	1.17	7.09(6.82, 7.37)	16.49(16.04, 16.95)
Ningnan	1.41	4.91(4.73, 5.10)	28.75(28.11, 29.36)	1.39	8.36 (8.01,8.73)	16.63(16.19, 17.08)	1.06	6.80(6.48, 7.13)	15.68(15.24, 16.12)
Puge	1.56	5.75(5.51, 5.99)	26.93(26.24, 27.64)	1.63	10.03(9.60, 10.49)	16.22(15.78, 16.66)	1.23	8.00(7.61, 8.40)	15.29(14.86, 15.73)
Butuo	1.65	5.73(5.52, 5.95)	28.75(28.11, 29.36)	1.72	10.30(9.85, 10.76)	16.63(16.19, 17.08)	1.4	8.89(8.47, 9.33)	15.68(15.24, 16.12)
Jinyang	1.69	5.76(5.55, 5.99)	29.18(28.54, 29.81)	1.71	9.99(9.56, 10.46)	17.06(16.63, 17.51)	1.43	8.89(8.46, 9.32)	16.08(15.65, 16.53)
Zhaojue	1.83	6.24(6.03, 6.45)	29.18(28.54, 29.81)	1.76	10.32(9.92, 10.73)	17.06(16.63, 17.51)	1.48	9.17(8.78, 9.56)	16.08(15.65, 16.53)
Xide	1.63	5.37(5.18, 5.56)	30.50(30.05, 30.95)	1.67	9.53(9.12, 9.94)	17.49(17.05, 17.95)	1.53	9.24(8.82, 9.68)	16.49(16.04, 16.95)
Mianning	1.44	5.00(4.85, 5.17)	28.75(28.11, 29.36)	1.39	8.35(8.05, 8.67)	16.63(16.19, 17.08)	1.09	6.99(6.71, 7.27)	15.68(15.24, 16.12)
Yuexi	1.54	5.27(5.10, 5.45)	29.18(28.54, 29.81)	1.55	9.07(8.72, 9.42)	17.06(16.63, 17.51)	1.29	8.04(7.72, 8.38)	16.08(15.65,1 6.53)
Ganluo	1.51	5.62(5.40, 5.84)	26.93(26.24, 27.64)	1.65	10.19(9.77, 10.62)	16.22(15.78, 16.66)	1.25	8.17(7.79, 8.54)	15.29(14.86, 15.73)
Meigu	1.71	6.34(6.10, 6.59)	26.93(26.24, 27.64)	1.69	10.38(9.96, 10.81)	16.22(15.78, 16.66)	1.21	7.91(7.54, 8.28)	15.29(14.86, 15.73)
Leibo	1.59	5.90(5.67, 6.12)	26.93(26.24, 27.64)	1.57	9.65(9.26, 10.06)	16.22(15.78, 16.66)	1.26	8.25(7.89, 8.63)	15.29(14.86, 15.73)
Liangshan	1.44	4.95(4.80, 5.11)	29.12(28.20, 30.07)	1.43	8.42(8.10, 8.75)	16.94(16.30, 17.61)	1.14	7.12(6.82, 7.42)	15.98(15.32, 16.67)

[a] Observed prevalence rate.
[b] Estimated prevalence rate.

3.3.3. Different numbers of clusters

Tables 3 to 5 illustrated the bias, log (MSE) and 95 % *PI* predicted by the model at a registration rate of 30 % (the true value of γ_1 was 0.7). When the number of clusters (K) was incorrectly specified, the predictions by the model exhibited increased bias and log (MSE). Notably, the bias and log (MSE) were more pronounced when K was set to 2 than when K was set to 6. This discrepancy was primarily attributed to the fact that when the K was smaller than the actual K, some areas would be incorrectly divided into the first cluster. As previously discussed, the prior distribution of the reporting rate for the first cluster of areas significantly influenced the model's performance. (See Table 4.)

3.3.4. Different numbers of counties in each cluster

According to the findings presented in Table 6, it could be observed that as the number of areas reporting the worst quality increased (Scenario 4), both the bias and log (MSE) of the model increased. This was likely due to the increased number of areas with the worst quality leaded to a higher underreporting rate in the simulated data. Conversely, when the number of areas reporting the best quality increased (Scenario 3), the model benefited from incorporating prior information for a larger number of areas, which resulted in relatively lower errors.

3.4. Recommendations based on the risk assessment

In underdeveloped areas where comprehensive capture-recapture sampling or field surveys are not feasible due to resource constraints, targeted investigations could be conducted based on the estimated underreporting risk. Areas identified with a high risk of underreporting could consider increasing the sample size for future investigations. For example, counties such as Muli, Yanyuan, Ganluo, Meigu, Leibo and Puge might benefit from this approach. Additionally, it was recommended to enhance the frequency of preventive and educational campaigns as part of routine operations. The underreporting risk for each county was shown in Fig. 8.

4. Discussion

This study aimed to estimate the risk of underreporting of NCDs in underdeveloped areas. By clustering the study areas based on the data reporting quality and introducing the underreporting rate in the areas with the best quality as a priori, this study utilized the CPM-CDQ proposed by Oliveira et al. [23] to estimate the potential prevalence and the underreporting risk of hypertension and diabetes, taking Liangshan Prefecture as an example. Sensitivity analyses were also performed to assess the predictive performance of the CPM-CDQ under different data conditions.

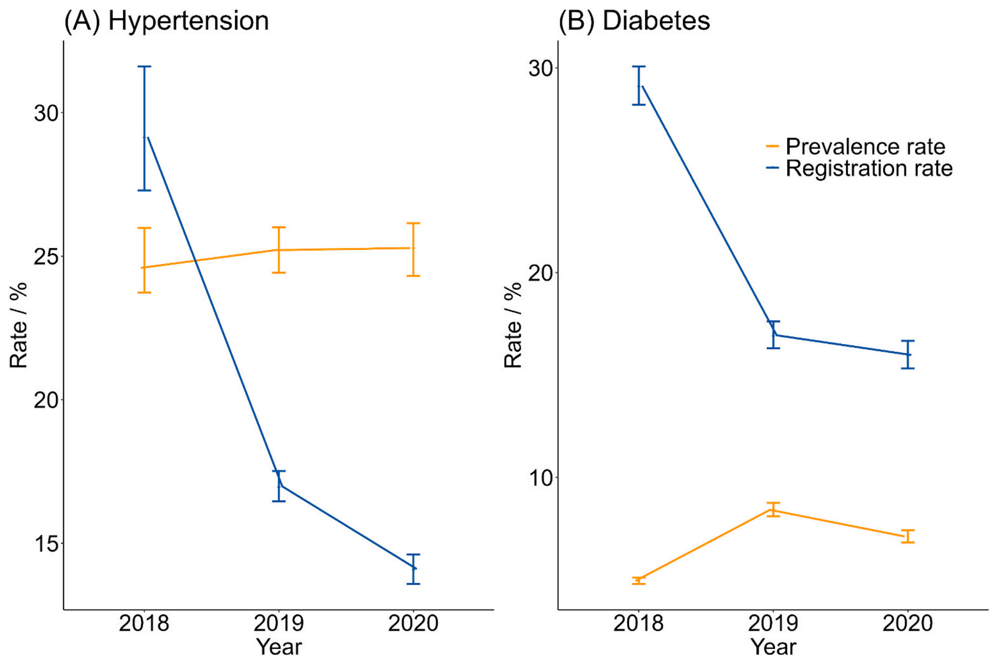

Fig. 5. Estimation of the prevalence and registration rate in Liangshan Prefecture from 2018 to 2020.

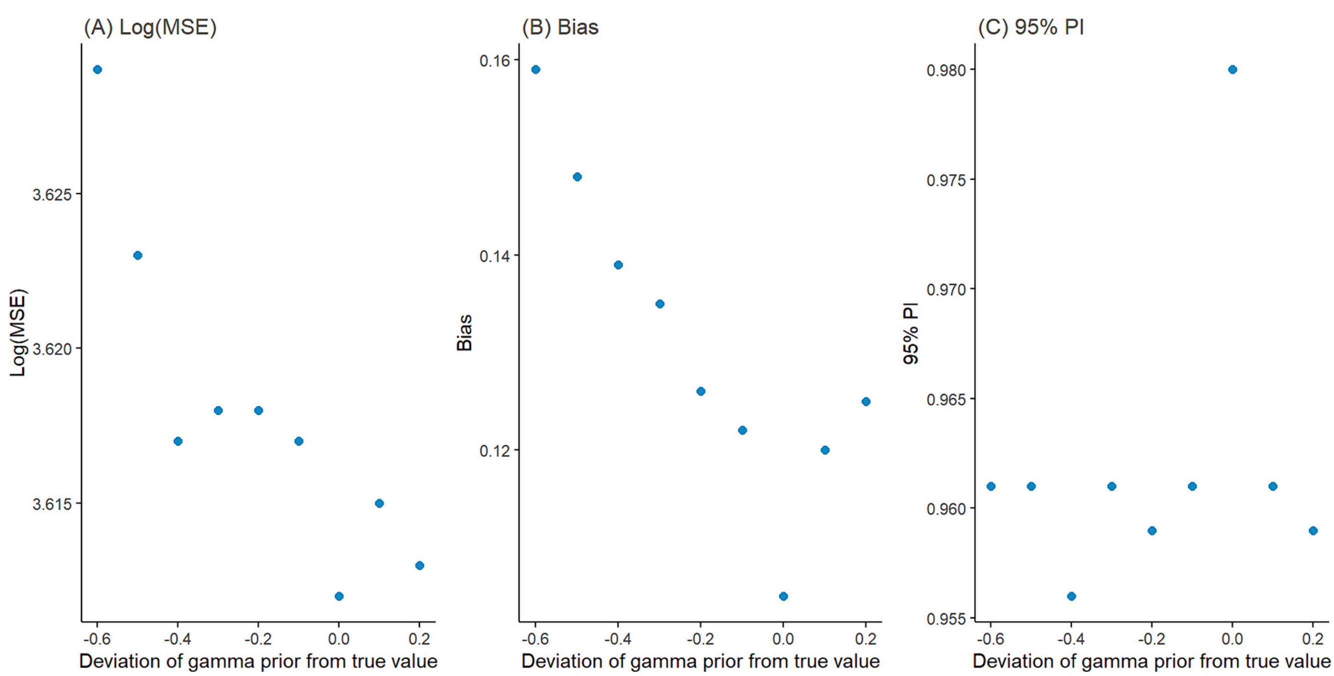

Fig. 6. Model performance under different levels of accuracy of prior information.

The findings highlighted a significant high prevalence of hypertension and diabetes in Liangshan Prefecture, accompanied by a concerningly low registration rate. The estimated results were close to those of earlier research conducted in Liangshan Prefecture. As an example, the estimated prevalence of hypertension in Xichang City in 2020 was 18.95 % (95 % *CI*: 18.34 %, 19.58 %), which was consistent with the results obtained through cross-sectional surveys [32]. In terms of registration rates, the overall registration rates for hypertensive and diabetes in Liangshan Prefecture were concerningly low, e.g., below the average in Sichuan Province [33]. Moreover, from 2018 to 2020, the registration rates for both hypertension and diabetes showed a declining trend. This likely resulted from the corrections of misinformation and false records initiated in 2019 and compounded by the impact of the COVID-19 pandemic, which led to delayed treatments for approximately 37.1 % of chronic disease patients [30,31].

Liangshan Prefecture is a multi-ethnic area, mainly inhabited by the Yi people. Yi residents normally exhibit a less balanced diet than Han residents [34], with an alcohol consumption rate of 60.1 %, which is notably higher than the Han residents. Given the established link between alcohol intake and hypertension risk [35,36], these lifestyle and dietary habits likely contribute to the elevated disease prevalence. The dietary habits of residents are also causes of the higher rate of diabetes in Liangshan Prefecture. The prevalence of diabetes among Chinese adults was 6.0 % in 2018, which was lower than the estimated diabetes

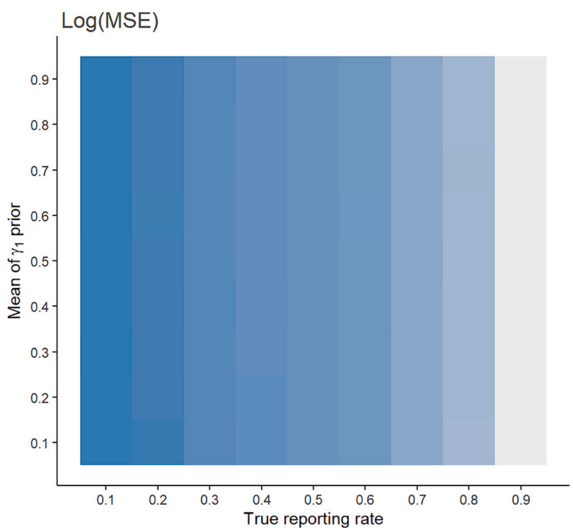

Fig. 7. Prediction error of the model under different levels of data completeness.

registration rate in 2018–2020 in Liangshan Prefecture [37]. The population in Liangshan Prefecture tends to consume high-fat and high-calorie foods, which leads to hypertension and obesity, both of which are risk factors for the development of diabetes [38,39].

This study demonstrated that the CPM-CDQ was flexible in assessing the potential risk of NCDs in underdeveloped areas and could be extended to a wide range of applications. In practical application, where a comprehensive overview of the study areas is challenging to acquire, preliminary surveys could normally be readily to be conducted to obtain data completeness information from some areas with the best data quality. This information could be served as a valuable priori knowledge for the model, enhancing its predictive accuracy. The sensitivity analyses have revealed that the prediction error of the model decreased as the registration rate increased. Consequently, when there are data with different integrities in practical application, prior information based on the data with a high registration rate could help to reduce the prediction error. In addition, the number of areas within each cluster could influence the model's performance. When the cluster with the best reporting quality encompassed a larger number of areas, the model's prediction error was reduced, as more areas contribute to the priori information. Therefore, in underdeveloped areas, it is advisable to establish a robust priori information for clusters with a substantial number of areas, contingent upon the data available. Subsequently, establishing inter-cluster dependencies could yield more precise estimates.

In this study, we applied a robust framework, the CPM-CDQ proposed by Oliveira et al. [23], for estimating the underreported risk of NCDs, particularly in underdeveloped areas. By focusing on hypertension and diabetes in Liangshan Yi Autonomous Prefecture, the findings provided valuable insights to inform health policy and resource allocation in areas with limited healthcare infrastructure. The key advantage of the CPM-CDQ is that it classifies and ranks areas based on registration quality, requiring registration rate of only areas with the best data quality, making it more widely applicable. While this study focused solely on hypertension and diabetes, the model is applicable to other non-communicable diseases due to its characteristic that it mainly utilized patient counting data. It enables an evaluation of the burden of chronic diseases and underreporting risk in areas where large-scale

Table 6
Prediction results of different models in different clusters of areas.

Different scenarios	Bias	log (MSE)	95 % *PI*
Scenario 1 m = 21, n = 10	0.982	3.622	0.961
Scenario 2 m = 10, n = 21	−0.306	3.615	0.980
Scenario 3 m = 26, n = 5	−0.205	3.568	0.980
Scenario 4 m = 5, n = 26	−1.300	3.821	0.960

Note: m represents the number of areas with the best quality; n represents the number of areas with the worst quality.

Table 3
Bias for prediction.

Number of clusters	Prior mean of γ_1(true value = 0.7)								
	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
$K = 4$ (true clusters)	0.114	0.108	0.117	0.113	0.111	0.122	0.105	0.120	0.125
$K = 2$	3.018	3.012	3.010	3.018	3.010	3.008	3.006	3.012	3.020
$K = 6$	0.804	0.799	0.810	0.780	0.813	0.809	0.791	0.809	0.798

Table 4
log(MSE) for prediction.

Number of clusters	Prior mean of γ_1 (true value = 0.7)								
	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
$K = 4$ (true clusters)	3.618	3.617	3.617	3.613	3.613	3.617	3.612	3.615	3.613
$K = 2$	4.149	4.149	4.146	4.145	4.148	4.146	4.144	4.147	4.150
$K = 6$	4.089	4.088	4.086	4.088	4.087	4.087	4.086	4.090	4.092

Table 5
95 % *PI* coverage for prediction.

Number of clusters	Prior mean of γ_1 (true value = 0.7)								
	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
$K = 4$ (true clusters)	0.961	0.961	0.978	0.978	0.978	0.961	0.980	0.967	0.961
$K = 2$	0.798	0.798	0.880	0.798	0.882	0.882	0.886	0.882	0.882
$K = 6$	0.840	0.843	0.842	0.843	0.843	0.843	0.845	0.843	0.843

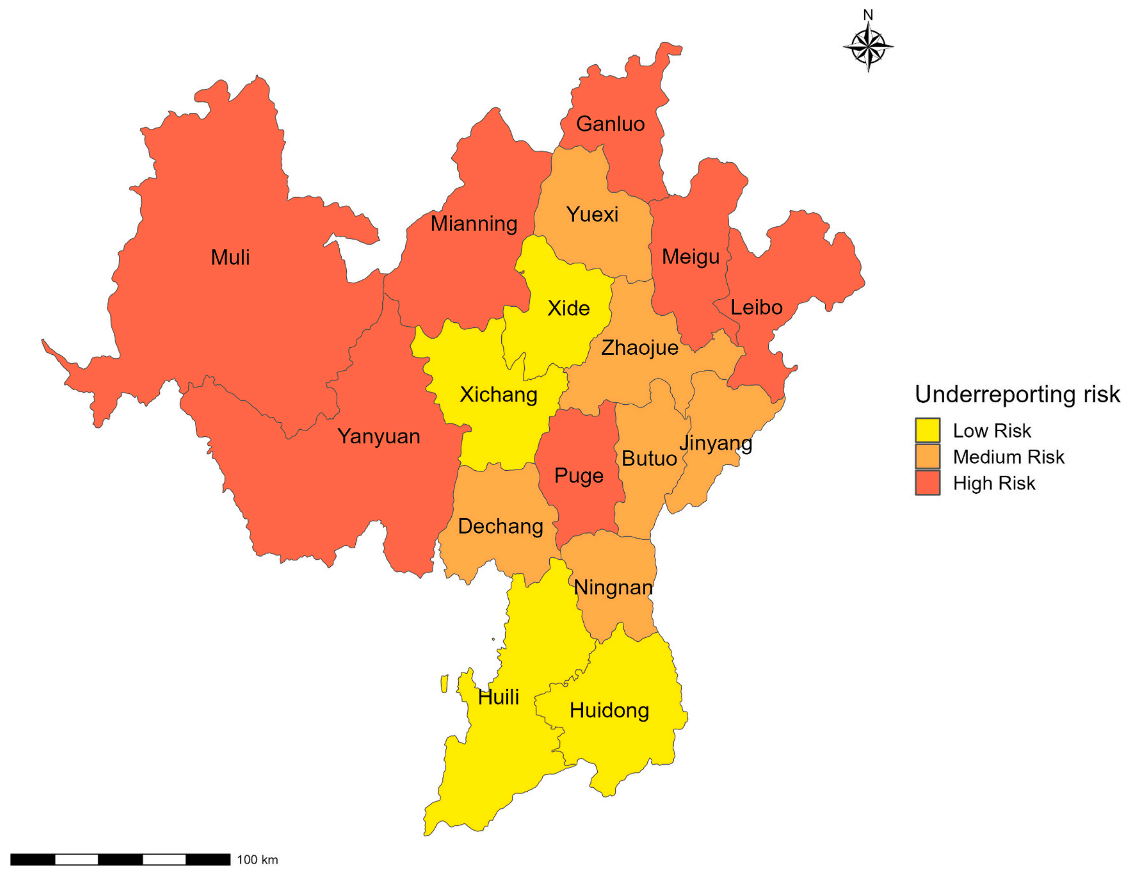

Fig. 8. Underreporting risk of Liangshan (The used base map was from http://bzdt.ch.mnr.gov.cn/).

surveys are difficult to carry out, thereby enhancing disease management [6,23]. Our application could provide a template for the application of the model in other underdeveloped areas.

However, the study acknowledged some limitations. First, the reliance on the reported data may introduce biases, particularly in areas with poor health infrastructure and reporting systems. Second, the causes of underreporting were multifaceted, and the study had only considered some objective indicators from a modeling perspective, while subjective factors, such as mindset and social culture, contributing to underreporting in underdeveloped areas should not be overlooked. Future research is encouraged to pay more attention to these subjective factors to improve the accuracy of estimating the risk of underreporting of NCDs in underdeveloped areas. Additionally, the present study only verified the application of the model in the correction of hypertension and diabetes underreporting, but according to the characteristics that the model mainly depends on counting data, the applicability of the model is not limited to this. Future research can further expand it to more public health problems (e.g., injuries) in more underdeveloped areas.

5. Conclusion

Liangshan Prefecture is experiencing a significant high prevalence of hypertension and diabetes, accompanied by a concerningly low registration rate. The CPM-CDQ model proved useful for assessing underreporting risks and facilitating targeted interventions for NCDs control and prevention, particularly in underdeveloped areas.

Funding

This work was supported by the Institute of New Productive Forces in Health, Sichuan Unviersity, Chengdu Municipal People's Government (grant H230413).

Ethics approval and consent to participate

This is an observational study. The data that support the findings of this study are available from the corresponding author upon reasonable request. No ethical approval or consent to participate is required.

Consent for publication

Not applicable.

CRediT authorship contribution statement

Hongli Wan: Conceptualization, Methodology, Software, Formal analysis, Data curation, Writing – original draft, Writing – review & editing. **Wenhui Zhu:** Conceptualization, Methodology, Software, Formal analysis, Data curation, Writing – original draft, Writing – review & editing. **Jingmin Yan:** Investigation, Writing – original draft. **Xinyue Han:** Investigation, Writing – original draft. **Jie Yu:** Investigation, Writing – original draft. **Qiang Liao:** Investigation, Resources, Writing – original draft. **Tao Zhang:** Conceptualization, Methodology, Formal analysis, Writing – original draft, Writing – review & editing, Supervision, Project administration, Funding acquisition.

Declaration of competing interest

The authors declare no competing interests.

Acknowledgements

The authors would like to thank Institute of New Productive Forces in

Health for financial support. The authors would like to thank the Liangshan Prefecture Center for Disease Control and Prevention, who helped us a lot in data collection.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.onehlt.2024.100889.

References

[1] WHO, Global Action Plan for the Prevention and Control of Noncommunicable Diseases 2013–2020. https://www.who.int/publications/i/item/9789241506236, 2013.

[2] WHO, The World Health Organization Report 2002: Reducing Risks, Promoting Healthy Life 2002. https://apps.who.int/iris/handle/10665/42510, 2020 (accessed 20 May 2020).

[3] Gojka Roglic, WHO global report on diabetes: a summary, IJNCD 1 (2016) 3–8, https://doi.org/10.4103/2468-8827.184853.

[4] WHO, Hypertension. https://www.who.int/news-room/fact-sheets/detail/hypertension, 2023 (accessed 16 March 2023).

[5] E. Ahmad, S. Lim, R. Lamptey, D.R. Webb, M.J. Davies, Type 2 diabetes, Lancet 400 (2022) 1803–1820, https://doi.org/10.1016/S0140-6736(22)01655-5.

[6] GBD 2021 Diabetes Collaborators, Global, regional, and national burden of diabetes from 1990 To 2021, with projections of prevalence to 2050: a systematic analysis for the global burden of disease study 2021, Lancet 402 (2023) 203–234, https://doi.org/10.1016/S0140-6736(23)01301-6.

[7] A. Budreviciute, S. Damiati, D.K. Sabir, K. Onder, P. Schuller-Goetzburg, G. Plakys, A. Katileviciute, S. Khoja, R. Kodzius, Management and prevention strategies for non-communicable diseases (NCDs) and their Risk factors, Front. Public Health 8 (2020) 574111, https://doi.org/10.3389/fpubh.2020.574111.

[8] American Diabetes Association Professional Practice Committee, 1. Improving care and promoting health in populations: standards of medical care in diabetes-2022, Diabetes Care 45 (2022) S8–S16, https://doi.org/10.2337/dc22-S001.

[9] American Diabetes Association, 4. Comprehensive Medical Evaluation and Assessment of Comorbidities: Standards of Medical Care in Diabetes-2020. Diabetes Care 43, 2020, pp. S37–S47, https://doi.org/10.2337/dc20-S004.

[10] The Royal Australian College of General Practitioners, Guidelines for Preventive Activities in General Practice. https://www.racgp.org.au/FSDEDEV/media/documents/Clinical%20Resources/Guidelines/Red%20Book/Guidelines-for-preventive-activities-in-general-practice.pdf, 2018 (accessed 2018).

[11] P. Geldsetzer, J. Manne-Goehler, M.E. Marcus, C. Ebert, Z. Zhumadilov, C. S. Wesseh, L. Tsabedze, A. Supiyev, L. Sturua, S.K. Bahendeka, A.M. Sibai, S. Quesnel-Crooks, B. Norov, K.J. Mwangi, O. Mwalim, R. Wong-McClure, M. T. Mayige, J.S. Martins, N. Lunet, D. Labadarios, L.M. Jaacks, The state of hypertension care in 44 low-income and middle-income countries: a cross-sectional study of nationally representative individual-level data from 1·1 million adults, Lancet 394 (2019) 652–662, https://doi.org/10.1016/S0140-6736(19)30955-9.

[12] NCD Risk Factor Collaboration (NCD-RisC), Long-term and recent trends in hypertension awareness, treatment, and control in 12 high-income countries: an analysis of 123 nationally representative surveys, Lancet 394 (2019) 639–651, https://doi.org/10.1016/S0140-6736(19)31145-6.

[13] K.L. Peterson, J.P. Jacobs, S. Allender, L.V. Alston, M. Nichols, Characterising the extent of misreporting of high blood pressure, high cholesterol, and diabetes using the Australian Health Survey, BMC Public Health 16 (2016) 695, https://doi.org/10.1186/s12889-016-3389-y.

[14] C. Pinto, R. Borrego, M. Eiró-Gomes, I. Casimiro, A. Raposo, T. Folha, D. Virella, A. C. Moreira, Embracing the nutritional assessment in cerebral palsy: a toolkit for healthcare professionals for daily practice, Nutrients 14 (2022) 1180, https://doi.org/10.3390/nu14061180.

[15] J. Sanyal, D. Rubin, I. Banerjee, A weakly supervised model for the automated detection of adverse events using clinical notes, J. Biomed. Inform. 126 (2022) 103969, https://doi.org/10.1016/j.jbi.2021.103969.

[16] L. Ongeri, D.A. Larsen, R. Jenkins, A. Shaw, H. Connolly, J. Lyon, S. Kariuki, B. Penninx, C.R. Newton, P. Sifuna, B. Ogutu, Community suicide rates and related factors within a surveillance platform in Western Kenya, BMC Psychiatr. 22 (2022) 7, https://doi.org/10.1186/s12888-021-03649-6.

[17] WHO, State of the World's Nursing 2020: Investing in Education, Jobs and Leadership 2020. https://www.who.int/publications/i/item/9789240003279, 2020 (accessed 6 April 2020).

[18] H. Wang, R. Liao, X. Chen, J. Yu, T. Zhu, Q. Liao, T. Zhang, How to improve the COVID-19 health education strategy in impoverished regions: a pilot study, Infect. Dis. Poverty 11 (2022) 38, https://doi.org/10.1186/s40249-022-00963-3.

[19] C. Guo, H. Wang, G. Feng, J. Li, C. Su, J. Zhang, Z. Wang, W. Du, B. Zhang, Spatiotemporal predictions of obesity prevalence in Chinese children and adolescents: based on analyses of obesogenic environmental variability and Bayesian model, Int. J. Obes. 43 (2019) 1380–1390, https://doi.org/10.1038/s41366-018-0301-0.

[20] M.O. Prates, Spatial extreme learning machines: an application on prediction of disease counts, Stat. Methods Med. Res. 28 (2019) 2583–2594, https://doi.org/10.1177/0962280218767985.

[21] S.B. Caudill, F.G. Mixon, Modeling household fertility decisions: estimation and testing of censored regression models for count data, Empir. Econ. 20 (1995) 183–196, https://doi.org/10.1007/BF01205434.

[22] O. Stoner, T. Economou, G. Drummond Marques da Silva, A hierarchical framework for correcting under-reporting in count data, J. Am. Stat. Assoc. 114 (2019) 1481–1492, https://doi.org/10.1080/01621459.2019.1573732.

[23] G.L. de Oliveira, R. Argiento, R.H. Loschi, R.M. Assunção, F. Ruggeri, M.D. Branco, E. Bias correction in clustered underreported data, Bayesian Anal. 17 (2022) 95–126, https://doi.org/10.1214/20-BA1244.

[24] X. Li, H.H. Chang, Q. Cheng, P.A. Collender, T. Li, J. He, L.A. Waller, B.A. Lopman, J.V. Remais, A spatial hierarchical model for integrating and bias-correcting data from passive and active disease surveillance systems, Spat. Spatiotemp. Epidemiol. 35 (2020) 100341, https://doi.org/10.1016/j.sste.2020.100341.

[25] R. Liao, L. Hu, Q. Liao, T. Zhu, H. Yang, T. Zhang, Analysis of death causes of residents in poverty-stricken areas in 2020: take Liangshan Yi autonomous prefecture in China as an example, BMC Public Health 22 (2022) 89, https://doi.org/10.1186/s12889-022-12504-6.

[26] L. Hu, W. Zhu, J. Yu, Y. Chen, J. Yan, Q. Liao, T. Zhang, Family-based improvement for health literacy among the Yi nationality (FAMILY) in Liangshan: protocol of an open cohort stepped wedge cluster randomized controlled trial, BMC Public Health 22 (2022) 1543, https://doi.org/10.1186/s12889-022-13782-w.

[27] Z. Yuan, Y. Wang, H. Zhang, Y. Wang, P. Lei, S. Wen, Y. Tang, The influence of dietary habits on blood pressure and blood glucose in Mianning County, Liangshan Yi Autonomous Prefecture, World Latest Med. Inform. 19 (2019) 198–199, in Chinese with English abstract, 10.19613/j.cnki.1671-3141.2019.17.145.

[28] M. Zhang, Y. Shi, B. Zhou, Z. Huang, Z. Zhao, C. Li, X. Zhang, G. Han, K. Peng, X. Li, Y. Wang, M. Ezzati, L. Wang, Y. Li, Prevalence, awareness, treatment, and control of hypertension in China, 2004-18: findings from six rounds of a national survey, BMJ 380 (2023) e071952, https://doi.org/10.1136/bmj-2022-071952.

[29] H. Song, D. Feng, R. Wang, J. Yang, Y. Li, J. Gao, Z. Wang, Z. Yan, C. Long, J. Zhou, Z. Feng, The urban-rural disparity in the prevalence and risk factors of hypertension among the elderly in China-a cross-sectional study, PeerJ 7 (2019) e8015, https://doi.org/10.7717/peerj.8015.

[30] W. Zheng, F. Kämpfen, Z. Huang, Health-seeking and diagnosis delay and its associated factors: a case study on COVID-19 infections in Shaanxi Province, China, Sci. Rep. 11 (2021) 17331, https://doi.org/10.1038/s41598-021-96888-2.

[31] T. Kendzerska, D.T. Zhu, A.S. Gershon, J.D. Edwards, C. Peixoto, R. Robillard, C. E. Kendall, The effects of the health system response to the COVID-19 pandemic on chronic disease management: a narrative review, Risk Manag. Healthc. Policy. 14 (2021) 575–584, https://doi.org/10.2147/RMHP.S293471.

[32] Y. Duandan, L. Qiang, Y. Fang, X. Hongli, C. Kai, Z. Tao, Analysis of prevalence of hypertension and related behaviors among residents based on community diagnosis in Xichang City 2021, J. Prevent. Med. Inform. 39 (2023) 304–311 (in Chinese with English abstract), https://link.cnki.net/urlid/51.1276.R.20221013.1456.001.

[33] Z. Sang, X. Guan, M. Wang, R. Mei, X. Tang, J. Liang, J. Zhang, X. Jiang, Sampling survey of basic public health service projects in Sichuan Province for the year 2014, Occup. Health Injury 31 (2016) 14–18 (in Chinese with English abstract).

[34] Y. Gao, X.W. Ran, X.H. Xie, H.L. Lu, T. Chen, Y. Ren, Y. Long, H.M. Tian, Prevalence of overweight and obesity among Chinese Yi nationality: a cross-sectional study, BMC Public Health 11 (2011) 919, https://doi.org/10.1186/1471-2458-11-919.

[35] F.D. Fuchs, S.C. Fuchs, The effect of alcohol on blood pressure and hypertension, Curr. Hypertens. Rep. 23 (2021) 42, https://doi.org/10.1007/s11906-021-01160-7.

[36] S. Ikehara, H. Iso, Alcohol consumption and risks of hypertension and cardiovascular disease in Japanese men and women, Hypertens. Res. 43 (2020) 477–481, https://doi.org/10.1038/s41440-020-0417-1.

[37] Y. Li, D. Teng, X. Shi, G. Qin, Y. Qin, H. Quan, B. Shi, H. Sun, J. Ba, B. Chen, J. Du, L. He, X. Lai, Y. Li, H. Chi, E. Liao, C. Liu, L. Liu, X. Tang, N. Tong, Z. Shan, Prevalence of diabetes recorded in mainland China using 2018 diagnostic criteria from the American Diabetes Association: national cross sectional study, BMJ 369 (2020) m997, https://doi.org/10.1136/bmj.m997.

[38] C. Wang, D. Wei, B. Wang, J. Zhang, K. Zhang, M. Ma, L. Pan, T. Yu, F. Xue, G. Shan, Effect of lifestyle on the prevalence of the metabolic syndrome among farmers, migrants with Yi ethnicity and Han population in Sichuan province of China, Asia Pac. J. Clin. Nutr. 19 (2010) 266–273. https://pubmed.ncbi.nlm.nih.gov/20460243/.

[39] J. Zhang, S. Wan, B. Zhang, F. Dong, L. Pan, W. Yihuo, H. Gong, F. Yang, G. Xu, Z. Li, G. Li, Y. Li, X. Wang, G. Shan, Twenty-year time trends in hypertension prevalence in Yi people of China: three successive cross-sectional studies, 1996-2015, BMJ Open 8 (2018) e022714, https://doi.org/10.1136/bmjopen-2018-022714.