Analysis of Influencing Factors of Community Management of Diabetes Based on Adaptive-Lasso Logistic Regression Model: A Community-based Study

Wei Lin
Chengdu University of Traditional Chinese Medicine
Yang Tian
Shizishan Community Health Service Center
Adeel Khoja
University of Adelaide
Xuan Zhao
Chengdu University of Traditional Chinese Medicine
Peng Hu
Chengdu University of Traditional Chinese Medicine
Mingyue Zheng (mingyue.zheng@adelaide.edu.au)
University of Adelaide

Research Article

Keywords: Community diabetes management, Adaptive-lasso logistic model, Risk management, Quality in health care, Health informatics

Posted Date: February 23rd, 2021

DOI: https://doi.org/10.21203/rs.3.rs-226156/v1

Analysis of Influencing Factors of Community Management of Diabetes based on Adaptive-Lasso Logistic Regression Model: A Community-based Study

Wei Lin[1], Yang Tian [2], Adeel Khoja[3], Xuan Zhao[4], Peng Hu[5], Mingyue Zheng[3,5*]

[1] School of Management, Chengdu University of Traditional Chinese Medicine, Chengdu, 611137, China;

[2] Shizishan Community Health Service Center, Chengdu, 610021, China;

[3] Adelaide Medical School, University of Adelaide, Adelaide, SA 5005, Australia;

[4] School of Pharmacy, Chengdu University of Traditional Chinese Medicine, Chengdu, 611137, China

[5] School of Health and Rehabilitation, Chengdu University of Traditional Chinese Medicine, Chengdu, 611137, China

Correspondence

Mingyue Zheng, Adelaide Medical School, University of Adelaide, Adelaide, SA 5005, Australia. School of Health and Rehabilitation, Chengdu University of Traditional Chinese Medicine, Chengdu, 611137, China. e-mail: mingyue.zheng@adelaide.edu.au

Abstract Background This study aimed to analyse which influencing factors may be more effective to achieve diabetes management targets in the community by the adaptive-lasso logistic regression model. **Methods** A cross-sectional study (N=1,127) was adopted to establish the adaptive-lasso logistic regression model of influencing factors for community management based on multi-stage cluster sampling data among patients with diabetes in China. Patient's fasting blood glucose level, blood pressure, and triglycerides was collected. **Results** Overall, 90.6% of included people had a fasting glucose level higher than 6.1mol/L, and 9.4% of them were below 6.1mol/L. By cross-validation, after folding eight times, the variables involved in the adaptive lasso-logistic regression model include age, education level, main source of income, marital status, average monthly income, free medical service, basic medical insurance for residents, hospital history, number of follow-up evaluations by family doctor team, voluntary participation in community blood glucose measurement. The Akaike Information Criterion and Bayesian Information Criterion of adaptive lasso-logistic regression model were 1980 and 2021, which were lower than the full-variable logistic model (2041, 2245) and the ridge logistic model (2043, 2348). The adaptive-lasso logistic regression model was better than the other two models regarding time cost. **Conclusions** The adaptive-lasso logistic regression model can analyse the influencing factors of community management in patients with diabetes. Community intervention and intensive management measures can significantly improve the blood glucose status of patients with diabetes.

Key words Community diabetes management, Adaptive-lasso logistic model, Risk management, Quality in health care, Health informatics

Background

The global costs of diabetes and its consequences are large and will substantially increase more than $2.1 trillion in 180 countries by 2030 [1], especially in low-and middle-income countries. Community-based diabetes management has become a cost-effective and cost-saving strategy for controlling and managing diabetes in primary care settings.[2-4] The community-based intervention positively reduced the HbA1c level of diabetic people and improved lipids and blood pressure control [5-8]. In primary care and community settings, interventions targeting diabetes management were better targeted at individuals with poor glycemic control, old age, and family history of diabetes and cardiovascular diseases [9-10]. Moreover, younger age and lower educational attainment were associated with lower probability of meeting the goals of diabetes management [11]. However, the influencing factors of diabetic patients based on community management are complex[5-9], and how to choose the most effective influencing factors requires in-depth research.

In terms of community health service, it proposed that patients' satisfaction with community health service was moderate, high satisfaction with the community health service shown better medication adherence and regular self-monitoring of blood glucose, and these associations varied by socioeconomic status[12, 13]. In particular, team-based care can improve patients' glucose levels, blood pressure, and lipid levels based on community preventive services task force [8] [14]. However, what effective services the community health service or the family doctor should provide to a diabetic patient is still to be investigated (i.e. consulation, education, free health service).

The adaptive-lasso logistic regression model was first proposed by Tibshirani (1996),[15] this model is to compress the parameters and make some regression coefficients gradually smaller or even close to zero. It has the advantages of subset selection and ridge regression. Huang (2008) proposed a new variable selection and estimation method based on the lasso model,

which can predict the correlation pattern between variables [16]. The lasso regression model has also been widely used in medical research for prediction and decision making [17-22]. Furthermore, many studies adopted the lasso model to screen variables and use them for optimization [23-25].

However, to the best our knowledge, no adaptive-lasso logistic regression model has been used to investigate the influencing factors of diabetes community management. Therefore, this study aimed to analyse which influencing factors may be more effective to achieve diabetes management targets (i.e. blood glucose, blood pressure, and triglycerides) in the community by the adaptive-lasso logistic regression model.

Methods

Data sources

A cross-sectional survey method (N=1,127) based on the multi-stage cluster sampling was used to survey selected samples from three communities from July 2019 to January 2021 (Figure 1). We surveyed the registered residents with the help from community health service centres in three communities in Sichuan Province. To avoid bias in the study design, we conducted a multidisciplinary expert demonstration and formulated reasonable inclusion and exclusion criteria to ensure higher reliability and validity when formulating the questionnaire. Random number tables were used to select registered diabetic patients in three communities in this study. The survey included demographic information (i.e. gender, age, marital status, education, number of children), self-management diabetes status (i.e. self-assessment of glycemic management, hospital history), basic situation of living (i.e. main source of income, average monthly income, medical payment methods), diabetes management in the community in the past six months (i.e. number of follow-up evaluations by the family doctor team, number of health consultations and free consultations, whether to participate in community blood glucose

measurement voluntarily), clinical measurements (i.e. latest fasting blood glucose level, blood pressure, and triglycerides), and other information. Family doctor follow-up records, blood glucose records, and health consultations and free consultations were derived from the residents' health management files and the Chengdu regional health information platform.

Participants whose home address or contact details were unchanged for more than three years were included in this study. The collected questionnaire (Supplementary Table 1) was double-entered by two researches to avoid bias in the data collection process. During the process of data collection and entry, there was no data loss.

Each participant signed an informed consent form, and they volunteered to withdraw from the study itself at any time without giving any reason. For illiterate and semi-illiterate participants, data were collected by investigators reading out the informed consent and questionnaires. Data collectors were not involved in the data analysis process to avoid bias. The Chronbach's α coefficient of the questionnaire is 0.927, the Kaiser-Meyer-Olkin value is 0.825, and the p-value of the spherical test is <0.001, indicating that the questionnaire has good reliability and validity in this study.

Lasso algorithm

Lasso algorithm is a regularisation method based on parameter estimation and variable selection. The parameter estimation is defined as follows:

$$\hat{\beta}_{\text{lasso}} = \arg\ \min{}^2 \| \mathrm{Y} - \sum_{j=1}^{p} X_j \beta_j \|^2 + \lambda \sum_{j=1}^{p} |\beta_j| \qquad (1-1)$$

In formula (1-1), λ is a regularised non-negative parameter, $\beta = (\beta_1, \beta_2, \ldots, \beta_p)^T$ is the regression coefficient, $X_j = (x_{1j}, x_{2j}, \ldots, x_{nj})^T$ and $j = 1, 2, \ldots, p$ are predictor variables, $X = (X_1, X_2, \ldots, X_p)^T$ is the predictor matrix, $\mathrm{Y} = (\mathrm{Y}_1, \mathrm{Y}_2, \ldots, \mathrm{Y}_n)^T$ is the response variable. $\lambda \sum_{j=1}^{p} |\beta_j|$ is called ' l penalty', when λ increases, the lasso model allows the

coefficients to approach zero; when $\lambda \to \infty$, the coefficients almost reach zero. The improvement of the adaptive-lasso method is to add different weights to different coefficients, and its expression is defined as:

$$\hat{\beta}^{*(n)} = \arg\ \min \|Y - \sum_{j=1}^{p} X_j \beta_j\|^2 + \lambda_n \sum_{j=1}^{p} \hat{w}_j |\beta_j| \qquad (1\text{-}2)$$

In formula (1-2), $\hat{w}_j = \frac{1}{|\hat{\beta}_j|^{\gamma}} (\gamma > 0)$, $j = 1, 2, \cdots, p$, where $\hat{\beta} = (\hat{\beta}_1, \hat{\beta}_2, \cdots, \hat{\beta}_p)^T$ is the coefficient estimated value obtained by the ordinary least squares method. Its weight vector expression is as follows:

$$\hat{W} = (\hat{w}_1, \hat{w}_2, \cdots, \hat{w}_p)^T = \left(\frac{1}{|\hat{\beta}_1|^{\gamma}}, \frac{1}{|\hat{\beta}_2|^{\gamma}}, \cdots, \frac{1}{|\hat{\beta}_p|^{\gamma}} \right) = \frac{1}{|\hat{\beta}|^{\gamma}} \quad (\gamma > 0) \qquad (1\text{-}3)$$

Adaptive-lasso estimation of logistic regression models

Based on logistic regression, $\hat{\beta}$ is defined as $\hat{\beta} = \arg\min^2 \left(-L(\beta) + \lambda \sum_{j=1}^{p} |\beta_j| \right)$, where $L(\beta)$ is its log-likelihood function. When the fasting glucose level of included patients is less than or equal to 6.1mol/L, y_i will be defined as 0; when the fasting glucose level is greater than 6.1mol/L, y_i will be defined as 1, then:

$$p_i(y_i) = p_i^{y_i}((1 - p_i)^{1 - y_i}) \quad i = 1, 2, \cdots, n$$

From this, we can obtain the log-likelihood function of the joint density function for n samples as follows: $L(\beta) = \sum_{i=1}^{n} p_i y_i = \sum_{i=1}^{n} \left(y_i \log \frac{p_i}{1 - p_i} + \log(1 - p_i) \right)$ (1-4)

Based on logistic regression, the estimated value of lasso can be expressed as follows:

$$\hat{\beta}_{\text{lasso}} = -\sum_{i=1}^{n} \left(y_i \left(\beta_{ij} + \sum_{i=1}^{p} \beta_i x_i \right) - \log\left(1 + \exp\left(\sum_{i=1}^{p} \beta_i x_i \right) \right) \right) + \lambda \sum_{j=1}^{p} |\beta_j| \qquad (1\text{-}5)$$

Statistical methods

Data were entered using Epidata (Version 3.1), data analysis was performed by R (Version 3.5.3), and Chi-square test and adaptive lasso-logistic regression were used to analyse the influencing factors of diabetes management. Due to the inconsistency of the dimensions, the data in this study is standardised using lar () in R, so that different eigenvalues have the same scale. Among them, the lasso-logistics regression model was completed by the glmnet package in R.

Model establishment and analysis

From July 2019 to January 2021, the questionnaire survey was conducted among diabetic patients in communities or primary health service centres in Chengdu City, Sichuan Province. In terms of sample size, one independent variable of the regression needs at least ten samples to support. A total of 17 independent variables were selected in this study. Considering the replacement of some samples in the sampling, elimination, and poor questionnaire compliance might lead to a reduction in the number of evaluable cases.

According to the statistics of the National Bureau of Statistics of China in 2018, the total permanent population of Chengdu is 16.3 million. The sample size estimation formula of large population is:

$$n = \frac{Z^2\sigma^2}{d} \qquad (1\text{-}6)$$

Where n is the total sample size, Z is the confidence interval, σ is the standard deviation, and d is the sampling error range. We choose Z=1.96, σ=0.5, and d=0.05. The minimum required sample size is determined to be 768. Therefore, a total of 1,200 questionnaires were distributed, and 1,127 valid questionnaires were collected. No participant withdrew from this study, but 73 of the surveys were considered invalid because they were incomplete. The effective response rate was 93.9%. The age range was 28-97 years. Regarding self-assessment of glycemic management, 32.5% of patients felt satisfied, 59.8% of patients felt generally satisfied, and 6.8%

of patients felt not very satisfied, and 6.9% of patients felt dissatisfied (Table 1).

Analysis of influencing factors for self-management in patients with diabetes

The univariate chi-square test was performed for each categorical variable (Table 1). It can be found that age, education level, main source of income, marital status, average monthly income, medical payment methods (basic medical insurance for residents, basic medical insurance for employee and free medical service), and hospital history were statistically significant ($p<0.05$).

Adaptive-lasso logistic regression analysis of influencing factors of diabetes self-management

Due to the inconsistency of the dimensions, the data in this study was standardised by using lar () in R, so that different feature values had the same scale. Lar () is known as least-angle regression, which is a reconstruction algorithm of adaptive-lasso logistic regression. This study adopted the latest fasting blood glucose measurement (0 or 1) as the dependent variable. The following factors that may affect the management of diabetes were assigned as independent variables (Supplementary Table 2). The logistic regression model of adaptive-lasso variable selection was introduced, and the coefficients of each selected variable were estimated to analyse the factors affecting diabetes management. The Lasso-logistics regression model of this study was completed by the glmnet package in R. The glmnet package fit a generalised linear model via penalised maximum likelihood. The relationship between the model error was obtained through cross-validation. The number of folding times was kept as eight . The selected variables and parameter estimates are shown in Table 2.

Comparing the three models in Table 2, it was found that the adaptive-lasso logistic model was more significant and concise in selecting variables. The Adaptive-lasso logistic model had the smallest Akaike Information Criterion (AIC) and Bayesian Information Criterion (BIC), and the full-variable logistic model had the largest AIC and BIC.

Comparison of model prediction accuracy

To predict the results of diabetes management, the established full-variable logistic model, adaptive-lasso logistic model and ridge logistic model were used. Table 3 shown the prediction accuracy of the models with different proportions of the training set. The prediction accuracy of the full-variable logistic model and the adaptive-lasso logistic model was slightly higher than that of the ridge logistic model (Table 3). However, the prediction accuracy of the three models were not significantly different, and the reason for the small difference in accuracy might be related to data collection. But only from the time required to collect data, the running time of the adaptive-lasso logistic model was less than the prediction time of the full-variable logistic model. When the prediction accuracy of the three models was close, or there was not much difference, the adaptive-lasso logistic model was better than the other two models from the perspective of time-cost.

RESULTS AND DISCUSSION

Community management of people with diabetes affected by multiple factors

Community diabetes management is complex and is affected by many variables/factors [5-10]. Adaptive-lasso logistic model, full-variable logistic model and ridge logistic model were used to screen variables, and then age, education, marital status, main source of income, monthly average income, basic medical insurance for employees, and whether they have been hospitalised were all included in the above three models. It indicated that these factors were related to self-management in patients with diabetes. Our findings indicated that the effects of age, marital status, and hospital history on self-management of patients with diabetes were significant, the result was consistent with Lian (2019) and Bansal (2018) [16,17,26,27]. It also found that younger diabetic patients have more stringent behaviours in blood glucose control. As age increases, the elderly need to reinforce their self-management of diabetes [23, 26]. Besides,

educational level was an influential factor in community management of diabetes, and higher the educational level, the better the self-glycemic control [11, 24]. In terms of public medical care, we also found that compared with those who did not receive public medical care, those who had public medical care had a slight advantage in blood glucose management. In addition, patients with a history of hospitalisation were better at blood glucose control than patients without a history of hospitalisation.

Adaptive-lasso logistic model can be used to analyse the influencing factors of diabetes community management

The parameter estimation of the adaptive-lasso logistic model was both unbiased and has the advantage of ridge regression and subset selection [23]. In this study, the adaptive-lasso algorithm was introduced into a logistic regression model to analyse the influencing factors among people with diabetes and to evaluate community diabetes management. According to the authors' knowledge, using this model to explore the influencing factors of blood glucose management from the perspective of community diabetes management has not been reported to date. The adaptive-lasso algorithm was used in the logistic regression model to achieve the purpose of filtering variables and simplifying the model in this study. Based on the cross-sectional survey data, comparing the fitted full-variable logistic model and the ridge logistic model, the AIC and BIC of the adaptive-lasso logistic model were the smallest; when making predictions, the time cost of the adaptive-lasso logistic model was the least. After analysis and comparison, the adaptive-lasso logistic model was more concise than the variables selected by full-variable logistic, and the model was more compressed. The adaptive-lasso logistic model considered variables that affect diabetes self-management including age, education level, the main source of income, marital status, average monthly income, free medical service, basic medical insurance for residents, hospital history, number of follow-up evaluations by family doctor team, voluntary participation in community blood glucose measurement.

Community intervention and intensive bio-markers measurements can significantly improve blood glucose management among patients with diabetes. In particular, increasing the number of follow-up evaluations, health consultations and free consultations by the family doctor team will significantly affect the blood glucose management effect indiabetic patients. Our results were also consistent with the results of Aponte (2017) [7]. It seems that practical strategies to improve diabetes self-management based on community include; social support for family members, health care and community members, and local free or low-cost diabetes education materials and courses [28].

However, this study also had a limitation, the survey of this study could not include lifestyle influencing factors such as physical activity and diet, considering the target population was based on the community setting.

Conclusions

The adaptive-lasso logistic model can be used in the analysis of diabetes community management factors. It can accurately screen out factors that affect diabetes management. The obtained model can better explain the indicators for these influencing factors and provide advice for primary care settings and communities. Diabetes management is affected by many factors, and relevant knowledge and education should be strengthened for elderly, non-married, less-educated, people with low-income, diabetic patients without public medical care and no previous hospitalisation. Moreover, the number of follow-up evaluations of family doctor teams and the frequency of health consultations and free consultations should increase to achieve optimal management of diabetes based on community.

List of abbrevations

AIC: Akaike Information Criterion; BIC: Bayesian Information Criterion.

Acknowledgements

All authors are very grateful for the support from the doctors and nurses of the community health service centres during the questionnaire distribution and data collection process from three communities in Chengdu City, Sichuan Province.

Author contributions

Data collection, YT; Formal analysis, WL; Funding acquisition, WL and MZ; Methodology, WL and MZ; Project administration, YT; Resources, XZ; Writing original draft, MZ and WL; Writing, reviewing and editing, AK, YT, PH, XZ and MZ.

Funding

This work was supported by the Primary Health Care Development Research Center of Sichuan Province (grant number SWFZ20-Y-032) and Chengdu University of Traditional Chinese Medicine (grant number ZRQN2019010) held by WL and (grant number RQN2019009) secured by MZ.

Ethics approval and consent to participate

This study was an observational cross-sectional study. The methods of the study followed the STROBE Statement. All methods were carried out in accordance with relevant guidelines and regulations. Ethics approval was obtained from the Human Ethics Committee of Chengdu University for Traditional Chinese Medicine (No. 2021KL-005). All participants signed an informed consent form, and they volunteered to withdraw from the questionnaire process at any time without giving any reason. For illiterate and semi-illiterate participants, data were collected by the investigators reading out the informed consent and the questionnaires. Data collectors were not involved in the data analysis process in order to avoid any bias.

Patient consent

Obtained

Patient and Public Involvement

Patients or the public were not involved in the design, or conduct, or reporting, or dissemination plans of our research.

Consent for publication

Not applicable

Data availability statement

All the data analysed as part of this study are included in the current manuscript.

Conflicts of interest

All the authors of the study stated that there are no conflicts of interest.

References

1. Bommer, C., et al., *Global Economic Burden of Diabetes in Adults: Projections From 2015 to 2030.* Diabetes Care, 2018. 41(5): p. 963-970.
2. Han, Hae-Ra, Siobhan McKenna, Manka Nkimbeng, Patty Wilson, Sally Rives, Olayinka Ajomagberin, Mohammad Alkawaldeh, Kelli Grunstra, Nisa Maruthur, and Phyllis Sharps. "*A systematic review of community health center based interventions for people with diabetes.*" Journal of community health. 2019.44(6): 1253-1280.
3. Zhou, Xilin, Karen R. Siegel, Boon Peng Ng, Shawn Jawanda, Krista K. Proia, Xuanping Zhang, Ann L. Albright, and Ping Zhang. "*Cost-effectiveness of diabetes prevention interventions targeting high-risk individuals and whole populations: a systematic review.*" Diabetes Care, 2020(43): 1593-1616.
4. Manne-Goehler, Jennifer, Pascal Geldsetzer, Kokou Agoudavi, Glennis Andall-Brereton, Krishna K. Aryal, Brice Wilfried Bicaba, Pascal Bovet et al. "*Health system performance for people with diabetes in 28 low-and middle-income countries: a cross-sectional study of nationally representative surveys*." PLoS medicine. 2019.16(3): e1002751.
5. Pamungkas, R.A. and K. Chamroonsawasdi, *HbA1c reduction and weight-loss outcomes: a systematic review and meta-analysis of community-based intervention trials among*

patients with type 2 diabetes mellitus. International Journal of Diabetes in Developing Countries, 2019. 39(2): p. 394-407

6. Little, T.V., et al., *Community health worker interventions for Latinos with type 2 diabetes: a systematic review of randomised controlled trials.* Curr Diab Rep, 2014. 14(12): p. 558
7. Aponte, J., et al., *Health effectiveness of community health workers as a diabetes self-management intervention.* Diab Vasc Dis Res, 2017. 14(4): p. 316-326.
8. Captieux M., et al., *Supported self-management for people with type 2 diabetes: a meta-review of quantitative systematic reviews*. BMJ open. 2018.1;8(12).
9. Murphy, M.E., et al., *Improving risk factor management for patients with poorly controlled type 2 diabetes: a systematic review of healthcare interventions in primary care and community settings.* BMJ Open, 2017. 7(8): p. e015135.
10. Omar, S.M., et al., *Prevalence, risk factors, and glycaemic control of type 2 diabetes mellitus in eastern Sudan: a community-based study.* Ther Adv Endocrinol Metab, 2019. 10: p. 2042018819860071.
11. Siegel, K.R., et al., *Prevalence of Major Behavioral Risk Factors for Type 2 Diabetes.* Diabetes Care, 2018. 41(5): p. 1032-1039.
12. Yin, T., et al., *Socioeconomic status moderates the association between patient satisfaction with community health service and self-management behaviors in patients with type 2 diabetes: A cross-sectional survey in China.* Medicine (Baltimore), 2019. 98(22): p. e15849.
13. Ong, Suan Ee, Joel Jun Kai Koh, Sue-Anne Ee Shiow Toh, Kee Seng Chia, Dina Balabanova, Martin McKee, Pablo Perel, and Helena Legido-Quigley. *"Assessing the influence of health systems on type 2 diabetes mellitus awareness, treatment, adherence, and control: a systematic review."* PloS one. 2018.3(13): e0195086.

14. *Community Preventive Services Task Force. Electronic address, y.c.g., Team-Based Care to Improve Type 2 Diabetes Management: Recommendation of the Community Preventive Services Task Force. Am J Prev Med, 2019. 57(1): p. e27-e29.*
15. Tibshirani R. *Regression shrinkage and selection via the lasso.* Journal of the Royal Statistical Society: Series B (Methodological). 1996;58(1):267-88.
16. Jian Huang, S. Ma, and C. Zhang, *Adaptive Lasso for Sparse High-dimensional Regression Midels.* Statistica Sinica 2008. 18, 1603-1618.
17. Liu K, Chen J, Zhang K, Wang S, Li X. A diagnostic prediction model of acute symptomatic portal vein thrombosis. Annals of vascular surgery. 2019 Nov 1;61:394-9.
18. Fatima, Nida, and Ashfaq Shuaib. "*Development and Validation of Machine Learning Algorithms for Predicting 30-Day Mortality Following Carotid Endarterectomy: Carotid Endarterectomy Mortality Scoring System (MMS)."* Neurosurgery. 2020.67(1). 47-394.
19. Backes, Yara, Matthijs P. Schwartz, Frank Ter Borg, Frank HJ Wolfhagen, John N. Groen, Wouter H. de Vos tot Nederveen, Jeroen van Bergeijk et al. *"Multicentre prospective evaluation of real-time optical diagnosis of T1 colorectal cancer in large non-pedunculated colorectal polyps using narrow band imaging (the OPTICAL study)." Gut* 2019.68. (2): 271-279.
20. Luo, Yan, Dongdong Mei, Jingshan Gong, Min Zuo, and Xiaojing Guo. *"Multiparametric MRI-based radiomics nomogram for predicting lymphovascular space invasion in endometrial carcinoma.*" Journal of Magnetic Resonance Imaging. 2020.4(52): 1257-1262.
21. Nwachukwu, Benedict U., Edward C. Beck, Elaine K. Lee, Jourdan M. Cancienne, Brian R. Waterman, Katlynn Paul, and Shane J. Nho. "Application of machine learning for predicting clinically meaningful outcome after arthroscopic femoroacetabular impingement surgery." *The American journal of sports medicine* 48, no. 2 (2020): 415-

423.

22. Murtojärvi, Mika, Anni S. Halkola, Antti Airola, Teemu D. Laajala, Tuomas Mirtti, Tero Aittokallio, and Tapio Pahikkala. "Cost-effective survival prediction for patients with advanced prostate cancer using clinical trial and real-world hospital registry datasets." *International journal of medical informatics* 133 (2020): 104014.
23. Sun M, Tian M. *A Class of Derivative-Free CG Projection Methods for Nonsmooth Equations with an Application to the LASSO Problem*. Bulletin of the Iranian Mathematical Society. 2019:1-23.
24. Shukor, S., et al., *Quantitative assessment of LASSO probe assembly and long-read multiplexed cloning.* BMC Biotechnol, 2019. 19(1): p. 50.
25. Jung, A. and N. Tran, *Localized Linear Regression in Networked Data.* 2019.
26. Lian, J., et al., *Long-term cost-effectiveness of a Patient Empowerment Programme for type 2 diabetes mellitus in primary care.* Diabetes Obes Metab, 2019. 21(1): p. 73-83.
27. Bansal, V., et al., *Inpatient diabetes management by specialised diabetes team versus primary service team in non-critical care units: impact on 30-day readmission rate and hospital cost.* BMJ Open Diabetes Res Care, 2018. 6(1): p. e000460.
28. Purnell, T.S., et al., *Perceived Barriers and Potential Strategies to Improve Self-Management Among Adults with Type 2 Diabetes: A Community-Engaged Research Approach.* Patient, 2016. 9(4): p. 349-58.

Figures

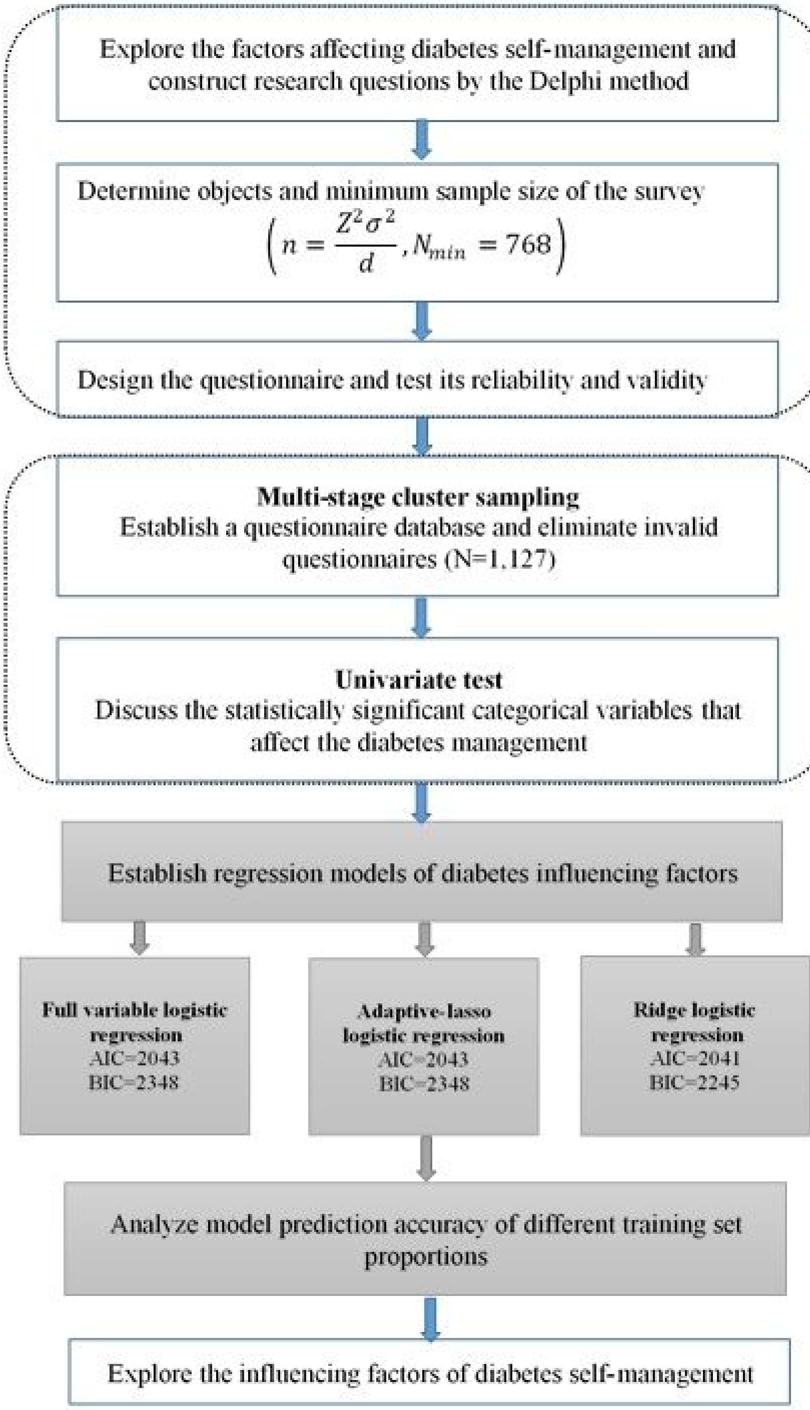

Figure 1

Flow chart of research design and implementation

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- SupplementaryTableAuthorsinformation.docx
- SupplementaryTable1.DiabetesCommunityQuestionnaire.docx
- SupplementaryTable2.VariableAssignment.docx