BMJ Open

BMJ Open is committed to open peer review. As part of this commitment we make the peer review history of every article we publish publicly available.

When an article is published we post the peer reviewers' comments and the authors' responses online. We also post the versions of the paper that were used during peer review. These are the versions that the peer review comments apply to.

The versions of the paper that follow are the versions that were submitted during the peer review process. They are not the versions of record or the final published versions. They should not be cited or distributed as the published version of this manuscript.

BMJ Open is an open access journal and the full, final, typeset and author-corrected version of record of the manuscript is available on our site with no access controls, subscription charges or pay-per-view fees (http://bmjopen.bmj.com).

If you have any questions on BMJ Open's open peer review process please email info.bmjopen@bmj.com

BMJ Open: first published as 10.1136/bmjopen-2019-035134 on 4 March 2020. Downloaded from http://bmjopen.bmj.com/ on April 20, 2024 by guest. Protected by copyright.

BMJ Open

Development of screening tools to predict the risk of recurrence and related complications following anal fistula surgery: a prospective cohort study

Journal:	*BMJ Open*
Manuscript ID	bmjopen-2019-035134
Article Type:	Protocol
Date Submitted by the Author:	19-Oct-2019
Complete List of Authors:	Mei, Zubing; Shuguang Hospital, Department of Anorectal Surgery Li, Yue; Shuguang Hospital, Department of Anorectal Surgery Zhang, Zhijun; Shuguang Hospital, Department of Anorectal Surgery Liu, Suzhi; Shuguang Hospital, Department of Anorectal Surgery Zhou, Haikun; Shuguang Hospital, Department of Anorectal Surgery Han, Ye; Shuguang Hospital, Department of Anorectal Surgery Du, Peixin; Shuguang Hospital, Department of Anorectal Surgery Shao, Zhuo; Changhai Hospital, Department of General Surgery Ge, Maojun; Shuguang Hospital, Department of General Surgery Wang, Qingming; Shuguang Hospital, Department of Anorectal Surgery Yang, Wei; Shuguang Hospital, Department of Anorectal Surgery
Keywords:	SURGERY, Risk management < HEALTH SERVICES ADMINISTRATION & MANAGEMENT, Colorectal surgery < SURGERY, Protocols & guidelines < HEALTH SERVICES ADMINISTRATION & MANAGEMENT

SCHOLARONE™
Manuscripts

BMJ Open: first published as 10.1136/bmjopen-2019-035134 on 4 March 2020. Downloaded from http://bmjopen.bmj.com/ on April 20, 2024 by guest. Protected by copyright.

I, the Submitting Author has the right to grant and does grant on behalf of all authors of the Work (as defined in the below author licence), an exclusive licence and/or a non-exclusive licence for contributions from authors who are: i) UK Crown employees; ii) where BMJ has agreed a CC-BY licence shall apply, and/or iii) in accordance with the terms applicable for US Federal Government officers or employees acting as part of their official duties; on a worldwide, perpetual, irrevocable, royalty-free basis to BMJ Publishing Group Ltd ("BMJ") its licensees and where the relevant Journal is co-owned by BMJ to the co-owners of the Journal, to publish the Work in this journal and any other BMJ products and to exploit all rights, as set out in our licence.

The Submitting Author accepts and understands that any supply made under these terms is made by BMJ to the Submitting Author unless you are acting as an employee on behalf of your employer or a postgraduate student of an affiliated institution which is paying any applicable article publishing charge ("APC") for Open Access articles. Where the Submitting Author wishes to make the Work available on an Open Access basis (and intends to pay the relevant APC), the terms of reuse of such Open Access shall be governed by a Creative Commons licence – details of these licences and which Creative Commons licence will apply to this Work are set out in our licence referred to above.

Other than as permitted in any relevant BMJ Author's Self Archiving Policies, I confirm this Work has not been accepted for publication elsewhere, is not being considered for publication elsewhere and does not duplicate material already published. I confirm all authors consent to publication of this Work and authorise the granting of this licence.

For peer review only

Development of screening tools to predict the risk of recurrence and related complications following anal fistula surgery: a prospective cohort study

Zubing Mei[1, 2]*; Yue Li[1]*; Zhijun Zhang[1]; Suzhi Liu[1]; Haikun Zhou[1]; Ye Han[1]; Peixin Du[1]; Zhuo Shao[3]; Maojun Ge[4]; Qingming Wang [1]; Wei Yang[1, 2]

*Contributed equally to this manuscript

1. Department of Anorectal Surgery, Shuguang Hospital, Shanghai University of Traditional Chinese Medicine, Shanghai, China;
2. Anorectal Disease Institute of Shuguang Hospital, Shanghai, China;
3. Department of General Surgery, Changhai Hospital, The Second Military Medical University, Shanghai, China;
4. Department of General Surgery, Shuguang Hospital, Shanghai University of Traditional Chinese Medicine, Shanghai, China.

Correspondence to:

Dr. Zubing Mei, MD, PhD

Department of Anorectal Surgery, Shuguang Hospital, Shanghai University of Traditional Chinese Medicine; Anorectal Disease Institute of Shuguang Hospital

528 Zhangheng Road, Shanghai 201203, China

Tel: 086-2120256187

E-mail: herrmayor@126.com

Professor. Wei Yang, MD

Department of Anorectal Surgery, Shuguang Hospital, Shanghai University of Traditional Chinese Medicine; Anorectal Disease Institute of Shuguang Hospital

528 Zhangheng Road, Shanghai 201203, China

Tel: 086-2120256187

E-mail: yangweiyishi@163.com

For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

Running head: Risk prediction models for AFR and related complications

Abstract Word Count: 298

Word Count (excluding abstract): 3,257

References: 53

Figures: 1

Key words: anal fistula, recurrence, prediction model, complication, surgery, cohort study

Funding statement: This work was supported by the National Natural Science Foundation of China (grant no. 81774112), Science and Technology Commission of Shanghai Municipality (grant no. 18401904600, 19401933900 and 16401971600), TCM dominant disease breeding project (grant no. zybz-2017018), the Plateau Peak Disciplines Innovation Team Project of Shanghai Municipal Education Commission (grant no. A1-N1920501010220).

Abbreviations: AF, anal fistula; AFR, anal fistula recurrence; BMI, body mass index; FCS, fully conditional specification; MVNI, multivariate normal imputation; TRIPOD, Transparent Reporting of a multivariable prediction model for Individual Prognosis Or Diagnosis.

Author contributions:

For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

Dr. Zubing Mei had full access to all of the data in the study and takes responsibility for the integrity of the data and the accuracy of the data analysis. Drs Mei and Li are co-first authors of this article.

Study concept and design: Zubing Mei.

Acquisition, analysis, or interpretation of data: Yue Li, Zubing Mei, Zhijun Zhang, Ye Han, Suzhi Liu, Haikun Zhou, Peixin Du, Qingming Wang, Wei Yang.

Drafting of the manuscript: Zubing Mei.

Critical revision of the manuscript for important intellectual content: All authors.

Statistical analysis: Zubing Mei, Zhuo Shao, Maojun Ge.

Administrative, technical, or material support: All authors.

Study supervision: Zubing Mei.

Competing interests statement:

The authors declare that they have no conflict of interest.

Role of the Funder/Sponsor:

The funder of the study had no role in the study design, data collection, data analysis, data interpretation or writing of the manuscript. The corresponding author had full access to all the data in the study and has final responsibility for the decision to submit for publication.

For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

Abstract

Introduction

Postoperative recurrence and related complications are common and related to poor outcome in patients with anal fistula. Besides the association with short and long−term cure rate, the perioperative complications have also been recently highlighted in these operated patients. This study aims to identify a set of predictive factors to develop risk prediction models for recurrence and related complications following anal fistula surgery. To accomplish this, we will apply a novel and comprehensive combination of patient-reported questionnaire instruments, psychophysical testing, laboratory and imaging findings to develop prediction models.

Methods and analysis

This is a prospective hospital-based cohort study using a linked database collected health data including Wechat questionnaires, laboratory and imaging findings, as well as follow-up outcomes for all adult patients who suffered from anal fistula at a tertiary referral hospital in Shanghai, China. We will construct logistic regression models to predict anal fistula recurrence (AFR) as well as related complications (eg, wound hemorrhage, edema, urinary retention, delayed wound healing and unplanned hospitalization) during and after AF surgery, and machine learning approaches will also applied to construct risk-prediction models. This prospective study is the first one investigating AFR and related complications using multi-dimensional variables. Due to the lack of effective means to monitor postoperative complications, prior

For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

prevention remains the best strategy. This study will provide alternative tools for early screening of high-risk patients of AFR and related complications, helping surgeons better understand the aetiology and outcome of anal fistula in an earlier stage.

Ethics and dissemination

The study is approved by the Institutional Review Board of Shuguang Hospital affiliated with Shanghai University of TCM (Approval Number: 2019-699-54-01). The results of this cohort will be submitted to international scientific peer-reviewed journals or conferences in surgery, anorectal surgery or anorectal diseases.

Trial registration number

Chinese Clinical Trial Registry (ChiCTR1900025069); Pre-results.

Key words

Anal fistula; treatment outcome; recurrence; surgery; cohort study; prediction model

For peer review only

For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

Strengths and limitations of this study

- This is a hospital-based prospective cohort study of patients with anal fistula at a tertiary referral hospital in China.
- Prediction models will be developed with a random sample of 60% of the AF cohort as the derivation cohort, and then validated with the remaining 40% as the validation cohort.
- Multidimensional clinically useful candidate predictors will be fully examined from a variety of sources including the published systematic review, Delphi surveys and univariable or multivariable logistic regression analysis.
- Bootstrapping procedure will be applied for the internal and external validation of the prediction models and multiple imputation will be used to treat the missing data.
- Non-response bias may occur as many of the variables are collected through a Wechat questioneer platform.
- Some more potentially predictors will not be involved or not collected in the current study, for example, data related to postoperative nursing stratergy and outpatient follow-up frequency.

review only

For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

Introduction

Anal fistula (AF) is a common perianal disease usually infected by cryptoglandular origin, which is regarded as a chronic stage of perianal abscess.[1] Postoperative recurrence, defined as persistence or recurrence of AF symptoms, or the development of recurrent perianal sepsis or chronic AF within six months of surgery, is one of the severe complications of AF surgery.[2 3] Our recently published meta-analysis based on 20 studies reported a recurrence rate of about 19% (95% CI 0.15-0.23) in patients having AF surgery.[4] Because of the high degree of difficulty of surgery for patients with high complex AF, the postoperative recurrence rate of these patients can be as high as 50%, and the failure rate of reoperation remains 10%.[5-8] It is considered as one of the most difficult and complicated diseases of anorectal department.

A large number of studies have shown that the recurrence of AF is related to multiple factors, such as unclear diagnosis, improper treatment of internal orifice and blind stump of fistula, omission and improper treatment of internal orifice, incorrect method of seton, omission of branch of the fistula and poor drainage.[9 10] Studies also reported that anal fistula recurrence (AFR) was associated with individual characteristics of the patients, such as history of the enteritis, previous anal surgery, obesity and smoking.[4 11-13] Li et.al retrospectively analyzed 1783 patients with AF receiving operation and found that the location of AF, operation history in the perianal region, seton history and enteritis were independent risk factors for recurrence of AF.[5]

For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

Mei et al. conducted a meta-analysis involving 20 studies with 6168 patients and concluded high transsphincteric fistula, internal opening unidentified, and horseshoe extensions were independent risk factors for AFR with high-quality evidence, while prior anal surgery, seton placement surgery, and multiple fistula tract were demonstrated as risk factors for AFR with moderate-quality evidence.[4] Factors influencing other perioperative complications related AF surgery including wound hemorrhage, edema, urinary retention, delayed wound healing and unplanned hospitalization are also rarely reported. Therefore, there is an urgent need to develop risk prediction tools for the complete profile of risk factors for AFR and those related complications.

Which AF patients will be cured after surgery and which ones will not, are rarely investigated. The development of a prediction model for AFR following surgery to identify patients with high risk, would be of significant importance. Firstly, surgeons can provide patients preoperatively about an estimated surgical cure rate according to the prediction models. Moreover, the current knowledge in the literature reporting potential predictive factors could instruct patients to avoid personal risk factors and adjust treatment strategy in order to improve the surgical cure rate, which have been well described and applied in the prevention of other diseases. [14-17]

However, so far, there are no effective screening tools to evaluate and predict the risk of recurrence or other adverse outcomes of AF. Therefore, the aim of the

For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

current study was to develop and validate multivariable prediction models that predict postoperative AFR and related complications.

Aims and objectives

The aim of this study is to develop risk prediction models for postoperative recurrence as well as other surgery-related complications in a prospective hospital-based AF cohort. Risk prediction model for perioperative complications including wound hemorrhage, edema, urinary retention, delayed wound healing and unplanned hospitalization will also be developed. Flowchart of prediction model development and assessment is provided in Figure 1.

The detailed tasks of this study are to:

1. Calculate the 3 to 6-month incidence of recurrence, and related complications in patients following AF surgery.

2. Establish the risk factors that significantly predict postoperative AFR and related complications of an AF cohort in a tertiary referral center.

3. Develop and validate risk prediction models for postoperative AFR and related complications of an AF cohort in a tertiary referral center.

We also have the following two hypotheses examined:

1. Patient-related demographic characteristics, fistula and surgery-related factors are predictive of postoperative AFR and related complications as dependent variables.

For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

2. The risk prediction models for postoperative AFR and related complications developed in our study have more than 70% of discriminating power.

Patients and Methods

Study design and participants

This study is a single-center, prospective observational study on a hospital-based cohort enrolled at a tertiary referral center in Shanghai, China.

Eligibility criteria

The enrollment of the cohort subject was initiated from June, 2019. All subjects who will undergo surgical intervention for AF will be included for inclusion. All operations will be performed by a group of colon and rectal surgeons at Shuguang Hospital, a regional tertiary referral center. Exclusion criteria are those whose age < 18 years, non-cryptoglandular fistula (eg, anal fistula due to inflammatory bowel disease, human immunodeficiency virus, malignant cancer, or obstetrical trauma), and rectovaginal or rectourethral fistula. The electronic medical records of the included subjects should be complete.

Trained clinical investigators are collecting data in several categories, including baseline demographics, laboratory examinations, surgical profiles, colonoscopic and MR imaging findings and postoperative outcomes within 3 to 6 months. Planned

For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

clinical reviews or electronic surveys are conducted during hospitalization and every half to 3 months after discharge for 6 months.

Data collection

The research team are all anorectal surgeons consisting of a principal investigator and supervised by the Ethics committee of Shuguang Hospital. Written informed consent was obtained from all patients. Investigators will not intervene in any aspects of patient surveys at every stage of follow-up. Data are collected using a convenient follow-up system supported by Empower EDC (OpenClinica, Boston, Massachussetts, USA). This electronic system introduces a machine learning algorithm, through which we can use the data already entered in the Empower system to train the algorithm model and let the system itself develop quality control algorithms, validate the entered data and identify missing or suspicious data. Finally, the data manager will check the missing or suspicious data, confirm their completeness and asked the data manager to provide additional data when necessary. Furthermore, an automatic reminder follow-up function also plays a pivotal role during the whole follow-up period.

Patient and public involvement

Patients and public will not be involved in the development, design, conduct or reporting of the study. The general results will be disseminated to participants through public education during follow-up.

For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

Clinical Outcomes

The primary study end point is postoperative recurrence following AF surgery defined as persistence or recurrence of AF symptoms, or the development of recurrent perianal sepsis or chronic AF within 3 to 6 months of surgery.[2 3 18] The second end point is a composite outcome of postoperative comorbidities or any of the following equivalent events including AFR, wound hemorrhage, edema, urinary retention, delayed wound healing or unplanned hospitalization associated with AF surgery. Outcomes were ascertained by the treating clinicians, medical records and interviews by the patients.

Selection of Predictor Variables

Candidate variables for the prediction model of the composite outcome of postoperative comorbidities will be screened according to the following pre-set criteria: (1) prior clinical knowledge; (2) results from a systematic review of the literature in April 2018 [4] with sufficient evidence to include them as predictive variables in the risk model for AFR as is demonstrated below; or (3) agreed upon by a group of anorectal surgeons or experts for their clinical relevance using a two-round Delphi survey. We initially identified the following covariates as relevant candidate variables based on our recent published systematic review and meta-analysis[4] as well as clinical knowledge and/or relevance. The determination of all other candidate variables are based on results of post-hoc analysis using univariable or multivariable survival analyses with a threshold of $p < 0.05$.

For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

Factors identified from our recent systematic review and Delphi survey (manuscript under review), will be measured at baseline. These include factors involving the identified significant risk factors which are reported in our meta-analysis are presented as follows:

► Prior anal surgery.

► Seton placement surgery.

► High transsphincteric fistula.

► Internal opening unidentified.

► Horseshoe extensions.

► Multiple fistula tracts.

Some of the demographic factors and surgery details will also be collected due to limited power in our literature review and some non-significant potential factors (eg, smoking or alcohol use) may be risk factors and are also included as follows:

► Gender.

► Age.

► Smoking.

► Alcohol use.

► Diabetes mellitus.

► Obesity.

► Preoperative seton drainage.

For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

►High internal opening.

►Postoperative drainage.

►Supralevator extensions.

Other factors like laboratory examinations and MR imaging parameters (height of the internal openings, height and number of fistula, etc.) will also be collected.

Categorization of Potential Predictors

For categorical predictors, we can code them as "factor" variables, with coding as dummy variables, for example, smoking is coded originally as "1" for never smoker, "2" for past smoker, and "3" for current smoker and never smoker was selected as the reference category. Similar manner can be applied with alcohol use.[19]

Continuous variables formally should be measured at an interval or ratio scale, and should be able to take any value in a range. We treat ordered variables as linear which is generally reasonable for prediction. In other cases, continuous predictors can be grouped with meaningful categorization; for example, body mass index (BMI) can be classified based on internationally recognized categories (i.e., underweight, normal weight, overweight, and obesity).[20] Based on previous experiences, we will be deriving some predictors based on the responses of the surveys. However, in case some subjectivity in the classifications of these predictors may occur, sensitivity analyses will be performed to examine the robustness of our definitions during model building and validation.

For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

Study quality control for the prediction models

Based on the summary of methodological quality and developmental stage of prediction models by van Oort et al., we are developing predesigned criteria for quality control of our prediction models, which can make us carry out the study and report the results more rigorously.[21 22] The methodological checklist of the study are presented in supplementary material.

Missing Data

Candidate predictors with more than 60% missingness will be excluded. For those less than 60% missingness, multiple imputation are to be performed by imputing 20 complete data sets using multivariate normal regression, [23-26] which can reasonably approximate the true distributional relationship between the missing values and the available ones.[27] Among various multiple imputation approaches, fully conditional specification (FCS) and multivariate normal imputation (MVNI) are preferred, because they have been proved to be generally less biased than complete-case analysis. They can both generate similar results in the presence of either binary or ordinal variables that are not generally normally distributed. [27]

Statistical analysis for model derivation

Logistic regression will be applied to develop our prediction models for the binary outcomes. All data processing and statistical analysis will be performed using

For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

EmpowerStats software (www. empowerstats.com; X&Y Solutions, Inc., Boston, MA, USA) and statistical software packages R (R Foundation, Vienna, Austria).

We will first study the association between each potential variable and the outcome based on univariable analysis. Variables are considered further for multivariable regression modelling when they are associated with a p-value less than 0.20. Normality or linearity will be evaluated for the continuous predictors. Fractional polynomials are advocated for associations between the continuous predictors and the outcome for non-linear relationships. [28 29] We will perform backward stepwise selection with a $p < 0.001$ as the inclusion threshold and a $p > 0.05$ as the exclusion threshold for each imputed data set.

Predictors which appear in the imputation models with an inclusion fraction of $\geqslant$50% are qualified for the final multivariable model. Though there is no consensus regarding the optimal method for selecting predictors for inclusion, backwards elimination is generally considered as the preferred procedure as reported by Mantel et al.[30] Forward stepwise procedure will also be performed to repeat the analysis to test the robustness of the models. Overall regression coefficient estimates of the models will be generated with the combination of the imputed datasets based on Rubin's Rules, while taking into account uncertainty in the imputed values.[23 24 26] Collinearity will also be assessed which refers to the fact that predictors can have strong correlation with each other, defined as correlation coefficient >0.8, or variance inflation factor >10.[31] Then we will examine the interactions among the regression models.

For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

Prediction model performance assessment

Prediction models will be developed with a random sample of 60% of the AF cohort as the derivation cohort, and then validated with the remaining sample of 40% of the cohort as the validation cohort. The predictive performance in the derivation and validation cohort will be evaluated and reported by examining measures of predictive accuracy, discrimination and calibration. Nagelkerke's R^2 and Brier score will be used for the measurement of predictive accuracy.[32 33] The discriminative ability of the prediction models are evaluated using several statistics, which are according to the discriminative and calibration ability in both derivation and validation AF cohort. Model discrimination means the ability of the models to differentiate between high-risk patients and low-risk patients (having high or low risk of AFR or surgery-related complications). This will be assessed via Harrell's concordance statistic (C-index).[34] The calculation of the C-index will be performed in each of the 20 imputed data sets, and then averaged based on Rubin's rule.[35] The model is interpreted as having no discriminatory ability when a value of C-index is 0.5, and has perfect discrimination when a value of C-index is 1.0.[34] Calibration implies the agreement between the predicted outcomes and the observed outcomes, which is evaluated with the Hosmer and Lemeshow test for goodness of fit in all imputed datasets presented with calibration plots.[36] Calibration-in-the-large, which defines as the agreement between mean observed outcomes and mean predictions, will also be assessed for calibration.[37]

For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

Internal and external validation of prediction model

To make the prediction models reproducible, we have to conduct internal validation. Bootstrapping technique, as one of the most attractive resampling techniques, is a mostly applied validation method, which seems to be most efficient for obtaining stable optimism-corrected estimates.[34 38] It has been reported that bootstrap validation is a feasible technique for most prediction models with at least a 500 bootstrap resampling procedure using Harrell's validate function, which can adjust the developed models for over-fitting.[39] We will also apply temporal validation as external validation using a more recent AF patient cohort. [40]

Sample size

Since there are no widely accepted methods for the estimation of the sample size requirements to derive and validate the risk prediction models, the size of this AF cohort will be calculated to have at least 10 events per candidate predictive variable, which will be expected to adequately power the logistic regression models.[41-43] We estimate that our center will be able to collect around 2000 cases and that the incidence of ARF would be around 5-20% for AFR and 20% for the composite outcome of postoperative comorbidities.[4 5] Recording at least 100 events would allow around 10 predictor variables to be entered into the model. Data will be collected for an estimated n=2000 participants (initial n=1200 for prediction model development and the next n=800 for internal validation of the derived model).

For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

Methodological quality control

Limited studies have identified specific criteria for quality control in a prediction model, but we have strictly adhered to the guidelines for the reporting of studies developing, validating multivariable clinical prediction models as is reported in the TRIPOD (Transparent Reporting of a multivariable prediction model for Individual Prognosis Or Diagnosis) Statement to ensure methodological rigour.[44] All issues have been addressed in this study in supplementary material.

Discussion

In this study, we plan to develop internally validated novel statistical models for the prediction of AFR as well as related complications (ie, wound hemorrhage, edema, urinary retention, delayed wound healing and unplanned hospitalization within 3 to 6 months after AF surgery) among AF patients. The models will be developed among a large AF cohort in a hospital-representative linked database with validated clinical information. The models are based on variables including Wechat questionnaires, laboratory and imaging findings, as well as follow-up outcomes which will be routinely collected at the time of enrollment.

According to the existing knowledge and published systematic review, it is highly plausible that a number of patient, fistula and surgery related characteristics (eg, patient-related variables such as gender, age, diabetes mellitus or obesity,[45-47] lifestyle

For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

factors such as smoking, alcohol abuse,[48 49] fistula-related factors such as number of fistula tracts, horseshoe extensions, classification and location of fistula, surgery-related variables such as prior anal surgery and postoperative drainage[48 50 51]) easily ascertainable before surgery may predict AFR. Similar risk prediction models exist in other diseases, such as the Framingham Risk Score model to predict cardiovascular disease risk[52] and the Korean Crohn's Disease Prediction (KCDP) model to predict the clinical course of Crohn's disease.[53] Until now, risk factor investigation of predictors of perioperative surgery-related complications has been limited to assessment of single predictors with small sample size. The risk prediction models can help inform surgeons regarding high risk AF patients based on the overall risk factors. The primary purpose of study was to develop two risk prediction models to facilitate surgeons in identifying AF patients who will have surgical treatment at higher risk of developing recurrence and surgery-related complications. The predictive models will help both clinicians and patients identify the risk of complications after AF surgery in advance, take necessary interventions to reduce the risk of surgery-related complications as well as the personal and social financial burden brough about by those complications. The accurate risk prediction models are especially instructive for the development of the optimal surgical plan to achieve optimal surgical outcome.

Our study has several strengths. To the best of our knowledge, this is the first study with the primary aim to develop, internally and externally validate multivariable

For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

prediction models for AFR and related complications following AF surgery. Multidimensional clinically useful candidate predictors will be fully examined from a variety of sources including our published systematic review, Delphi surveys and univariable or multivariable logistic regression analysis. Second, we will apply the internal and external validation of the prediction models using bootstrapping procedure. Third, multiple imputation will also be used to treat the missing data. Last but not least, our study is a prospective cohort one with adequate follow-up period which can minimize certain bias.

Our study also has limitations. As many of the variables are collected through a Wechat questioneer platform during the hospitalization and follow-up, non-response bias can occur. To solve this issue, we reguarly send reminders to those who does not respond after discharge. Second, though we will investigate a series of potential predictors, some more potentially predictors will not be involved or not collected in the current study, data related to postoperative nursing stratergy and outpatient follow-up frequency for example.

The newly developed risk algorithms may have significant applications in clinical practice by helping recommend optimal surgical approach for a specific AF patient, as well as intensive perioperative care and education, timely assessment and discussion of the need for interventions to those most at high risk of developing recurrence or surgery-related complications. The models will specifically identify AF

For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

patients who are likely to develop recurrence or related complications following AF surgery to offer the quantitative evaluation of the risk. Moreover, the models will also provide reference information for preventing recurrence and reducing the rate of recurrence after operation, and to intervene some high risk factors in the early stage.

CONCLUSION

This study protocol summarizes the design of development and validation studies for a risk screening tool in patients receiving AF surgery. Results from this study will be interpreted for the purpose of clinical decision making. The models to be developed of the study could be used to make new recommendations for perioperative AF patients.

For peer review only

For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

References

1. Tantiphlachiva K. 7 - Anorectal Abscess and Fistula. In: Coss-Adame E, Remes-Troche JM, eds. Anorectal Disorders: Academic Press 2019:81-100.

2. Abbas MA, Jackson CH, Haigh PI. Predictors of outcome for anal fistula surgery. Archives of surgery (Chicago, Ill : 1920) 2011;146(9):1011-6. doi: 10.1001/archsurg.2011.197

3. Hamadani A, Haigh P, Liu I, et al. Who is at risk for developing chronic anal fistula or recurrent anal sepsis after initial perianal abscess? Dis Colon Rectum 2009;52(2):217-21.

4. Mei Z, Wang Q, Zhang Y, et al. Risk Factors for Recurrence after anal fistula surgery: A meta-analysis. International Journal of Surgery 2019;69:153-64. doi: https://doi.org/10.1016/j.ijsu.2019.08.003

5. Li J, Yang W, Huang Z, et al. [Clinical characteristics and risk factors for recurrence of anal fistula patients]. Zhonghua wei chang wai ke za zhi = Chinese journal of gastrointestinal surgery 2016;19(12):1370-74.

6. Gottgens KW, Smeets RR, Stassen LP, et al. Systematic review and meta-analysis of surgical interventions for high cryptoglandular perianal fistula. International journal of colorectal disease 2015;30(5):583-93. doi: 10.1007/s00384-014-2091-8

For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

7. Gottgens KW, Janssen PT, Heemskerk J, et al. Long-term outcome of low perianal fistulas treated by fistulotomy: a multicenter study. International journal of colorectal disease 2015;30(2):213-9. doi: 10.1007/s00384-014-2072-y

8. Bakhtawar N, Usman M. Factors Increasing the Risk of Recurrence in Fistula-in-ano. Cureus 2019;11(3):e4200. doi: 10.7759/cureus.4200

9. Sygut A, Mik M, Trzcinski R, et al. How the location of the internal opening of anal fistulas affect the treatment results of primary transsphincteric fistulas. Langenbeck's archives of surgery 2010;395(8):1055-9. doi: 10.1007/s00423-009-0562-0

10. Dziki A, Bartos M. Seton treatment of anal fistula: experience with a new modification. The European journal of surgery = Acta chirurgica 1998;164(7):543-8. doi: 10.1080/110241598750005930

11. Hermann J, Eder P, Banasiewicz T, et al. Current management of anal fistulas in Crohn's disease. Przeglad gastroenterologiczny 2015;10(2):83-8. doi: 10.5114/pg.2015.49684

12. Hamadani A, Haigh PI, Liu IL, et al. Who is at risk for developing chronic anal fistula or recurrent anal sepsis after initial perianal abscess? Diseases of the colon and rectum 2009;52(2):217-21. doi: 10.1007/DCR.0b013e31819a5c52

For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

13. Ellis CN, Clark S. Effect of tobacco smoking on advancement flap repair of complex anal fistulas. Diseases of the colon and rectum 2007;50(4):459-63. doi: 10.1007/s10350-006-0829-2

14. Moons KG, Altman DG, Vergouwe Y, et al. Prognosis and prognostic research: application and impact of prognostic models in clinical practice. Bmj 2009;338:b606. doi: 10.1136/bmj.b606

15. Ming C, Viassolo V, Probst-Hensch N, et al. Machine learning techniques for personalized breast cancer risk prediction: comparison with the BCRAT and BOADICEA models. Breast cancer research : BCR 2019;21(1):75. doi: 10.1186/s13058-019-1158-4

16. Khan SS, Ning H, Shah SJ, et al. 10-Year Risk Equations for Incident Heart Failure in the General Population. Journal of the American College of Cardiology 2019;73(19):2388-97. doi: 10.1016/j.jacc.2019.02.057

17. Terry MB, Liao Y, Whittemore AS, et al. 10-year performance of four models of breast cancer risk: a validation study. The Lancet Oncology 2019;20(4):504-17. doi: 10.1016/S1470-2045(18)30902-1

18. Yano T, Asano M, Matsuda Y, et al. Prognostic factors for recurrence following the initial drainage of an anorectal abscess. International journal of colorectal disease 2010;25(12):1495-8. doi: 10.1007/s00384-010-1011-9 [published Online First: 2010/07/20]

For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

19. Steyerberg EW. Coding of Categorical and Continuous Predictors. In: Steyerberg EW, ed. Clinical Prediction Models: A Practical Approach to Development, Validation, and Updating. Cham: Springer International Publishing 2019:175-90.

20. Morgan TM, Elashoff RM. Effect of Categorizing a Continuous Covariate on the Comparison of Survival Time. Journal of the American Statistical Association 1986;81(396):917-21. doi: 10.1080/01621459.1986.10478350

21. van Oort L, van den Berg T, Koes BW, et al. Preliminary state of development of prediction models for primary care physical therapy: a systematic review. Journal of Clinical Epidemiology 2012;65(12):1257-66. doi: https://doi.org/10.1016/j.jclinepi.2012.05.007

22. Rushton AB, Verra ML, Emms A, et al. Development and validation of two clinical prediction models to inform clinical decision-making for lumbar spinal fusion surgery for degenerative disorders and rehabilitation following surgery: protocol for a prospective observational study. BMJ Open 2018;8(5):e021078. doi: 10.1136/bmjopen-2017-021078

23. Wood AM, White IR, Royston P. How should variable selection be performed with multiply imputed data? Statistics in Medicine 2008;27(17):3227-46. doi: 10.1002/sim.3177

For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

24. White IR, Royston P, Wood AM. Multiple imputation using chained equations: Issues and guidance for practice. Statistics in Medicine 2011;30(4):377-99. doi: 10.1002/sim.4067

25. Schafer JL. Multiple imputation: a primer. Statistical Methods in Medical Research 1999;8(1):3-15. doi: 10.1177/096228029900800102

26. Marshall A, Altman DG, Holder RL, et al. Combining estimates of interest in prognostic modelling studies after multiple imputation: current practice and guidelines. BMC Medical Research Methodology 2009;9(1):57. doi: 10.1186/1471-2288-9-57

27. Lee KJ, Carlin JB. Multiple Imputation for Missing Data: Fully Conditional Specification Versus Multivariate Normal Imputation. American Journal of Epidemiology 2010;171(5):624-32. doi: 10.1093/aje/kwp425

28. Sauerbrei W, Royston P, Binder H. Selection of important variables and determination of functional form for continuous predictors in multivariable model building. Statistics in Medicine 2007;26(30):5512-28. doi: 10.1002/sim.3148

29. Royston P, Sauerbrei W. Building multivariable regression models with continuous covariates in clinical epidemiology--with an emphasis on fractional polynomials. Methods of information in medicine 2005;44(4):561-71.

30. Mantel N. Why Stepdown Procedures in Variable Selection. Technometrics 1970;12(3):621-25. doi: 10.1080/00401706.1970.10488701

For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

31. Vittinghoff E, Glidden DV, Shiboski SC, et al. Regression methods in biostatistics: linear, logistic, survival, and repeated measures models: Springer Science & Business Media 2011.

32. Steyerberg EW, Vickers AJ, Cook NR, et al. Assessing the performance of prediction models: a framework for traditional and novel measures. Epidemiology 2010;21(1):128-38. doi: 10.1097/EDE.0b013e3181c30fb2

33. NAGELKERKE NJD. A note on a general definition of the coefficient of determination. Biometrika 1991;78(3):691-92. doi: 10.1093/biomet/78.3.691

34. Harrell Jr FE. Regression modeling strategies: with applications to linear models, logistic and ordinal regression, and survival analysis: Springer 2015.

35. Rubin DB. Multiple imputation for nonresponse in surveys: John Wiley & Sons 2004.

36. Hosmer Jr DW, Lemeshow S, Sturdivant RX. Applied logistic regression: John Wiley & Sons 2013.

37. Snell KI, Ensor J, Debray TP, et al. Meta-analysis of prediction model performance across multiple studies: Which scale helps ensure between-study normality for the C-statistic and calibration measures? Statistical methods in medical research 2018;27(11):3505-22.

For peer review only

38. Chatfield C. Model uncertainty, data mining and statistical inference. Journal of the Royal Statistical Society: Series A (Statistics in Society) 1995;158(3):419-44.

39. Steyerberg EW, Bleeker SE, Moll HA, et al. Internal and external validation of predictive models: a simulation study of bias and precision in small samples. Journal of clinical epidemiology 2003;56(5):441-47.

40. Justice AC, Covinsky KE, Berlin JA. Assessing the generalizability of prognostic information. Annals of internal medicine 1999;130(6):515-24.

41. Concato J, Peduzzi P, Holford TR, et al. Importance of events per independent variable in proportional hazards analysis I. Background, goals, and general strategy. Journal of clinical epidemiology 1995;48(12):1495-501.

42. Peduzzi P, Concato J, Kemper E, et al. A simulation study of the number of events per variable in logistic regression analysis. Journal of clinical epidemiology 1996;49(12):1373-79.

43. Steyerberg EW. Clinical prediction models: A practical approach to development, validation, and updating Springer 2009.

44. Moons KG, Altman DG, Reitsma JB, et al. Transparent Reporting of a multivariable prediction model for Individual Prognosis or Diagnosis (TRIPOD): explanation and elaboration. Annals of internal medicine 2015;162(1):W1-W73.

For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

45. Hamadani A, Haigh PI, Liu I-LA, et al. Who is at risk for developing chronic anal fistula or recurrent anal sepsis after initial perianal abscess? Diseases of the colon & rectum 2009;52(2):217-21.

46. Sonoda T, Hull T, Piedmonte MR, et al. Outcomes of primary repair of anorectal and rectovaginal fistulas using the endorectal advancement flap. Diseases of the colon & rectum 2002;45(12):1622-28.

47. Hyman N, O'Brien S, Osler T. Outcomes after fistulotomy: results of a prospective, multicenter regional study. Diseases of the Colon & Rectum 2009;52(12):2022-27.

48. Ellis CN, Clark S. Effect of tobacco smoking on advancement flap repair of complex anal fistulas. Diseases of the colon & rectum 2007;50(4):459-63.

49. Schwandner T, Roblick MH, Kierer W, et al. Surgical treatment of complex anal fistulas with the anal fistula plug: a prospective, multicenter study. Diseases of the Colon & Rectum 2009;52(9):1578-83.

50. Garcia-Aguilar J, Belmonte C, Wong WD, et al. Anal fistula surgery: Factors associated with recurrence and incontinence. Diseases of the colon & rectum 1996;39(7):723-29.

51. Athanasiadis S, Köhler A, Nafe M. Treatment of high anal fistulae by primary occlusion of the internal ostium, drainage of the intersphincteric space, and mucosal advancement flap. International journal of colorectal disease 1994;9(3):153-57.

For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

52. Fox ER, Samdarshi TE, Musani SK, et al. Development and validation of risk prediction models for cardiovascular events in black adults: the Jackson Heart Study Cohort. JAMA cardiology 2016;1(1):15-25.

53. Park Y, Cheon JH, Park YL, et al. Development of a novel predictive model for the clinical course of Crohn's disease: results from the connect study. Inflammatory bowel diseases 2017;23(7):1071-79.

For peer review only

Figure legend

Figure 1. Flowchart of prediction model development and assessment.

For peer review only

For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

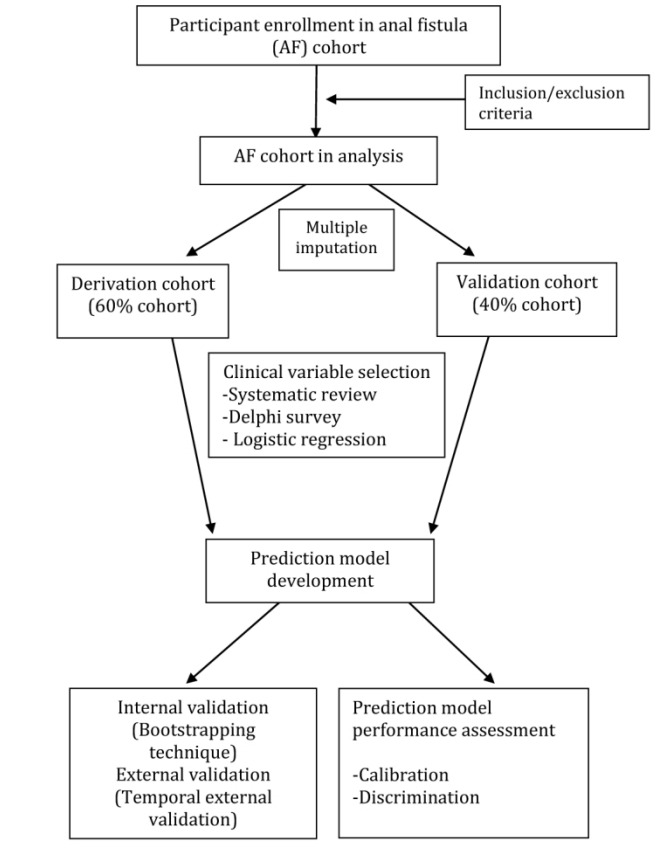

Figure 1. Flowchart of prediction model development and assessment.

BMJ Open: first published as 10.1136/bmjopen-2019-035134 on 4 March 2020. Downloaded from http://bmjopen.bmj.com/ on April 20, 2024 by guest. Protected by copyright.

TRIPOD Checklist: Prediction Model Development and Validation

This document contains a completed TRIPOD checklist for the manuscript: "**Development of screening tools to predict the risk of recurrence and related complications following anal fistula surgery: a prospective cohort study"**.

Zubing Mei, Yue Li, Zhijun Zhang, Suzhi Liu, Haikun Zhou, Ye Han, Peixin Du, Zhuo Shao, Maojun Ge, Qingming Wang, Wei Yang.

As this TRIPOD checklist refers to a study protocol, not all items are relevant at this stage. We have tried to the furthest extent possible to make the protocol adhere to the TRIPOD checklist and all reporting of results will be in accordance to the protocol and the TRIPOD statement. Page numbers in the submitted manuscript are provided. For items that are only partly relevant at this time, page numbers are provided in parentheses and for items that are not relevant at this time a "-" has been written.

Section/Topic	Item		Checklist Item	Page
Title and abstract				
Title	1	D;V	Identify the study as developing and/or validating a multivariable prediction model, the target population, and the outcome to be predicted.	1
Abstract	2	D;V	Provide a summary of objectives, study design, setting, participants, sample size, predictors, outcome, statistical analysis, results, and conclusions.	4-5
Introduction				
Background and objectives	3a	D;V	Explain the medical context (including whether diagnostic or prognostic) and rationale for developing or validating the multivariable prediction model, including references to existing models.	7-8
	3b	D;V	Specify the objectives, including whether the study describes the development or validation of the model or both.	7-9
Methods				
Source of data	4a	D;V	Describe the study design or source of data (e.g., randomized trial, cohort, or registry data), separately for the development and validation data sets, if applicable.	10
	4b	D;V	Specify the key study dates, including start of accrual; end of accrual; and, if applicable, end of follow-up.	10-11
Participants	5a	D;V	Specify key elements of the study setting (e.g., primary care, secondary care, general population) including number and location of centres.	11-12
	5b	D;V	Describe eligibility criteria for participants.	10-11
	5c	D;V	Give details of treatments received, if relevant.	10
Outcome	6a	D;V	Clearly define the outcome that is predicted by the prediction model, including how and when assessed.	12
	6b	D;V	Report any actions to blind assessment of the outcome to be predicted.	12
Predictors	7a	D;V	Clearly define all predictors used in developing or validating the multivariable prediction model, including how and when they were measured.	12-14
	7b	D;V	Report any actions to blind assessment of predictors for the outcome and other predictors.	14
Sample size	8	D;V	Explain how the study size was arrived at.	18
Missing data	9	D;V	Describe how missing data were handled (e.g., complete-case analysis, single imputation, multiple imputation) with details of any imputation method.	15
Statistical analysis methods	10a	D	Describe how predictors were handled in the analyses.	15-16
	10b	D	Specify type of model, all model-building procedures (including any predictor selection), and method for internal validation.	15-18
	10c	V	For validation, describe how the predictions were calculated.	18
	10d	D;V	Specify all measures used to assess model performance and, if relevant, to compare multiple models.	17
	10e	V	Describe any model updating (e.g., recalibration) arising from the validation, if done.	-
Risk groups	11	D;V	Provide details on how risk groups were created, if done.	-
Development vs. validation	12	V	For validation, identify any differences from the development data in setting, eligibility criteria, outcome, and predictors.	17-18
Results				
Participants	13a	D;V	Describe the flow of participants through the study, including the number of participants with and without the outcome and, if applicable, a summary of the follow-up time. A diagram may be helpful.	Figure 1
	13b	D;V	Describe the characteristics of the participants (basic demographics, clinical features, available predictors), including the number of participants with missing data for predictors and outcome.	-
	13c	V	For validation, show a comparison with the development data of the distribution of important variables (demographics, predictors and outcome).	-
Model development	14a	D	Specify the number of participants and outcome events in each analysis.	-
	14b	D	If done, report the unadjusted association between each candidate predictor and outcome.	-
Model	15a	D	Present the full prediction model to allow predictions for individuals (i.e., all regression	-

For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

TRIPOD Checklist: Prediction Model Development and Validation

specification			coefficients, and model intercept or baseline survival at a given time point).	-
	15b	D	Explain how to the use the prediction model.	-
Model performance	16	D;V	Report performance measures (with CIs) for the prediction model.	-
Model-updating	17	V	If done, report the results from any model updating (i.e., model specification, model performance).	-
Discussion				
Limitations	18	D;V	Discuss any limitations of the study (such as nonrepresentative sample, few events per predictor, missing data).	21
Interpretation	19a	V	For validation, discuss the results with reference to performance in the development data, and any other validation data.	-
	19b	D;V	Give an overall interpretation of the results, considering objectives, limitations, results from similar studies, and other relevant evidence.	-
Implications	20	D;V	Discuss the potential clinical use of the model and implications for future research.	21-22
Other information				
Supplementary information	21	D;V	Provide information about the availability of supplementary resources, such as study protocol, Web calculator, and data sets.	-
Funding	22	D;V	Give the source of funding and the role of the funders for the present study.	2

*Items relevant only to the development of a prediction model are denoted by D, items relating solely to a validation of a prediction model are denoted by V, and items relating to both are denoted D;V. We recommend using the TRIPOD Checklist in conjunction with the TRIPOD Explanation and Elaboration document.

For peer review only

TRIPOD Checklist: Prediction Model Development and Validation

This document contains a completed TRIPOD checklist for the manuscript: "**Development of screening tools to predict the risk of recurrence and related complications following anal fistula surgery: a prospective cohort study"**.

Zubing Mei, Yue Li, Zhijun Zhang, Suzhi Liu, Haikun Zhou, Ye Han, Peixin Du, Zhuo Shao, Maojun Ge, Qingming Wang, Wei Yang.

As this TRIPOD checklist refers to a study protocol, not all items are relevant at this stage. We have tried to the furthest extent possible to make the protocol adhere to the TRIPOD checklist and all reporting of results will be in accordance to the protocol and the TRIPOD statement. Page numbers in the submitted manuscript are provided. For items that are only partly relevant at this time, page numbers are provided in parentheses and for items that are not relevant at this time a "-" has been written.

Section/Topic	Item		Checklist Item	Page
Title and abstract				
Title	1	D;V	Identify the study as developing and/or validating a multivariable prediction model, the target population, and the outcome to be predicted.	1
Abstract	2	D;V	Provide a summary of objectives, study design, setting, participants, sample size, predictors, outcome, statistical analysis, results, and conclusions.	4-5
Introduction				
Background and objectives	3a	D;V	Explain the medical context (including whether diagnostic or prognostic) and rationale for developing or validating the multivariable prediction model, including references to existing models.	7-8
	3b	D;V	Specify the objectives, including whether the study describes the development or validation of the model or both.	7-9
Methods				
Source of data	4a	D;V	Describe the study design or source of data (e.g., randomized trial, cohort, or registry data), separately for the development and validation data sets, if applicable.	10
	4b	D;V	Specify the key study dates, including start of accrual; end of accrual; and, if applicable, end of follow-up.	10-11
Participants	5a	D;V	Specify key elements of the study setting (e.g., primary care, secondary care, general population) including number and location of centres.	11-12
	5b	D;V	Describe eligibility criteria for participants.	10-11
	5c	D;V	Give details of treatments received, if relevant.	10
Outcome	6a	D;V	Clearly define the outcome that is predicted by the prediction model, including how and when assessed.	12
	6b	D;V	Report any actions to blind assessment of the outcome to be predicted.	12
Predictors	7a	D;V	Clearly define all predictors used in developing or validating the multivariable prediction model, including how and when they were measured.	12-14
	7b	D;V	Report any actions to blind assessment of predictors for the outcome and other predictors.	14
Sample size	8	D;V	Explain how the study size was arrived at.	18
Missing data	9	D;V	Describe how missing data were handled (e.g., complete-case analysis, single imputation, multiple imputation) with details of any imputation method.	15
Statistical analysis methods	10a	D	Describe how predictors were handled in the analyses.	15-16
	10b	D	Specify type of model, all model-building procedures (including any predictor selection), and method for internal validation.	15-18
	10c	V	For validation, describe how the predictions were calculated.	18
	10d	D;V	Specify all measures used to assess model performance and, if relevant, to compare multiple models.	17
	10e	V	Describe any model updating (e.g., recalibration) arising from the validation, if done.	-
Risk groups	11	D;V	Provide details on how risk groups were created, if done.	-
Development vs. validation	12	V	For validation, identify any differences from the development data in setting, eligibility criteria, outcome, and predictors.	17-18
Results				
Participants	13a	D;V	Describe the flow of participants through the study, including the number of participants with and without the outcome and, if applicable, a summary of the follow-up time. A diagram may be helpful.	Figure 1
	13b	D;V	Describe the characteristics of the participants (basic demographics, clinical features, available predictors), including the number of participants with missing data for predictors and outcome.	-
	13c	V	For validation, show a comparison with the development data of the distribution of important variables (demographics, predictors and outcome).	-
Model development	14a	D	Specify the number of participants and outcome events in each analysis.	-
	14b	D	If done, report the unadjusted association between each candidate predictor and outcome.	-
Model	15a	D	Present the full prediction model to allow predictions for individuals (i.e., all regression	-

For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

TRIPOD Checklist: Prediction Model Development and Validation

specification			coefficients, and model intercept or baseline survival at a given time point).	-
	15b	D	Explain how to the use the prediction model.	-
Model performance	16	D;V	Report performance measures (with CIs) for the prediction model.	-
Model-updating	17	V	If done, report the results from any model updating (i.e., model specification, model performance).	-
Discussion				
Limitations	18	D;V	Discuss any limitations of the study (such as nonrepresentative sample, few events per predictor, missing data).	21
Interpretation	19a	V	For validation, discuss the results with reference to performance in the development data, and any other validation data.	-
	19b	D;V	Give an overall interpretation of the results, considering objectives, limitations, results from similar studies, and other relevant evidence.	-
Implications	20	D;V	Discuss the potential clinical use of the model and implications for future research.	21-22
Other information				
Supplementary information	21	D;V	Provide information about the availability of supplementary resources, such as study protocol, Web calculator, and data sets.	-
Funding	22	D;V	Give the source of funding and the role of the funders for the present study.	2

*Items relevant only to the development of a prediction model are denoted by D, items relating solely to a validation of a prediction model are denoted by V, and items relating to both are denoted D;V. We recommend using the TRIPOD Checklist in conjunction with the TRIPOD Explanation and Elaboration document.

For peer review only

BMJ Open

Development of screening tools to predict the risk of recurrence and related complications following anal fistula surgery: protocol for a prospective cohort study

Journal:	*BMJ Open*
Manuscript ID	bmjopen-2019-035134.R1
Article Type:	Protocol
Date Submitted by the Author:	17-Dec-2019
Complete List of Authors:	Mei, Zubing; Shuguang Hospital, Department of Anorectal Surgery Li, Yue; Shuguang Hospital, Department of Anorectal Surgery Zhang, Zhijun; Shuguang Hospital, Department of Anorectal Surgery Zhou, Haikun; Shuguang Hospital, Department of Anorectal Surgery Liu, Suzhi; Shuguang Hospital, Department of Anorectal Surgery Han, Ye; Shuguang Hospital, Department of Anorectal Surgery Du, Peixin; Shuguang Hospital, Department of Anorectal Surgery Qin, Xiufang; Shuguang Hospital, Department of Nursing Shao, Zhuo; Changhai Hospital, Department of General Surgery Ge, Maojun; Shuguang Hospital, Department of General Surgery Wang, Qingming; Shuguang Hospital, Department of Anorectal Surgery Yang, Wei; Shuguang Hospital, Department of Anorectal Surgery
Primary Subject Heading:	Surgery
Secondary Subject Heading:	Medical management
Keywords:	SURGERY, Risk management < HEALTH SERVICES ADMINISTRATION & MANAGEMENT, Colorectal surgery < SURGERY, Protocols & guidelines < HEALTH SERVICES ADMINISTRATION & MANAGEMENT

SCHOLARONE™
Manuscripts

BMJ Open: first published as 10.1136/bmjopen-2019-035134 on 4 March 2020. Downloaded from http://bmjopen.bmj.com/ on April 20, 2024 by guest. Protected by copyright.

I, the Submitting Author has the right to grant and does grant on behalf of all authors of the Work (as defined in the below author licence), an exclusive licence and/or a non-exclusive licence for contributions from authors who are: i) UK Crown employees; ii) where BMJ has agreed a CC-BY licence shall apply, and/or iii) in accordance with the terms applicable for US Federal Government officers or employees acting as part of their official duties; on a worldwide, perpetual, irrevocable, royalty-free basis to BMJ Publishing Group Ltd ("BMJ") its licensees and where the relevant Journal is co-owned by BMJ to the co-owners of the Journal, to publish the Work in this journal and any other BMJ products and to exploit all rights, as set out in our licence.

The Submitting Author accepts and understands that any supply made under these terms is made by BMJ to the Submitting Author unless you are acting as an employee on behalf of your employer or a postgraduate student of an affiliated institution which is paying any applicable article publishing charge ("APC") for Open Access articles. Where the Submitting Author wishes to make the Work available on an Open Access basis (and intends to pay the relevant APC), the terms of reuse of such Open Access shall be governed by a Creative Commons licence – details of these licences and which Creative Commons licence will apply to this Work are set out in our licence referred to above.

Other than as permitted in any relevant BMJ Author's Self Archiving Policies, I confirm this Work has not been accepted for publication elsewhere, is not being considered for publication elsewhere and does not duplicate material already published. I confirm all authors consent to publication of this Work and authorise the granting of this licence.

For peer review only

Development of screening tools to predict the risk of recurrence and related complications following anal fistula surgery: protocol for a prospective cohort study

Zubing Mei[1, 2]*; Yue Li[1]*; Zhijun Zhang[1]; Haikun Zhou[1]; Suzhi Liu[1]; Ye Han[1]; Peixin Du[1]; Xiufang Qin[3], Zhuo Shao[4]; Maojun Ge[5]; Qingming Wang [1]; Wei Yang[1, 2]

*Contributed equally to this manuscript

1. Department of Anorectal Surgery, Shuguang Hospital, Shanghai University of Traditional Chinese Medicine, Shanghai, China;
2. Anorectal Disease Institute of Shuguang Hospital, Shanghai, China;
3. Department of Nursing, Shuguang Hospital, Shanghai University of Traditional Chinese Medicine, Shanghai, China;
4. Department of General Surgery, Changhai Hospital, The Second Military Medical University, Shanghai, China;
5. Department of General Surgery, Shuguang Hospital, Shanghai University of Traditional Chinese Medicine, Shanghai, China.

Correspondence to:

Dr. Zubing Mei, MD, PhD

Department of Anorectal Surgery, Shuguang Hospital, Shanghai University of Traditional Chinese Medicine; Anorectal Disease Institute of Shuguang Hospital

528 Zhangheng Road, Shanghai 201203, China

Tel: 086-2120256187

E-mail: herrmayor@126.com

Professor. Wei Yang, MD

Department of Anorectal Surgery, Shuguang Hospital, Shanghai University of Traditional Chinese Medicine; Anorectal Disease Institute of Shuguang Hospital

528 Zhangheng Road, Shanghai 201203, China

For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

Tel: 086-2120256187

E-mail: yangweiyishi@163.com

Running head: Risk prediction models for AFR and related complications

Abstract Word Count: 300

Word Count (excluding abstract): 3,709

References: 53

Figures: 1

Key words: anal fistula, recurrence, prediction model, complication, surgery, cohort study

Funding statement: This work was supported by the National Natural Science Foundation of China (grant no. 81774112, 81673768), Science and Technology Commission of Shanghai Municipality (grant no. 18401904600, 19401933900 and 16401971600), TCM dominant disease breeding project (grant no. zybz-2017018), the Plateau Peak Disciplines Innovation Team Project of Shanghai Municipal Education Commission (grant no. A1-N1920501010220), Shanghai Clinical Key Specialty (Chinese Medicine Anorectal Disease) Construction Foundation of the 13th Five-Year Plan (grant no. shslczdzk04302), Shanghai Famous Old Traditional Chinese Medicine Experts Inheritance Studio Construction Program (grant no.SHGZS-2017028), Famous Old Traditional Chinese Medicine Experts Inheritance Studio Construction Program of Shanghai University of TCM (grant no.SZYMZYGZS4016-2017028).

For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

For peer review only

Abbreviations: AF, anal fistula; AFR, anal fistula recurrence; BMI, body mass index; FCS, fully conditional specification; MVNI, multivariate normal imputation; TRIPOD, Transparent Reporting of a multivariable prediction model for Individual Prognosis Or Diagnosis.

Author contributions:

Dr. Zubing Mei had full access to all of the data in the study and takes responsibility for the integrity of the data and the accuracy of the data analysis. Drs Mei and Li are co-first authors of this article.

Study concept and design: Zubing Mei.

Acquisition, analysis, or interpretation of data: Yue Li, Zubing Mei, Zhijun Zhang, Ye Han, Suzhi Liu, Haikun Zhou, Peixin Du, Xiufang Qin, Qingming Wang, Wei Yang.

Drafting of the manuscript: Zubing Mei.

Critical revision of the manuscript for important intellectual content: All authors.

Statistical analysis: Zubing Mei, Zhuo Shao, Maojun Ge.

Administrative, technical, or material support: All authors.

Study supervision: Zubing Mei.

Competing interests statement:

The authors declare that they have no conflict of interest.

For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

Role of the Funder/Sponsor:

The funder of the study had no role in the study design, data collection, data analysis, data interpretation or writing of the manuscript. The corresponding author had full access to all the data in the study and has final responsibility for the decision to submit for publication.

For peer review only

For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

Abstract

Introduction

Postoperative recurrence and related complications are common and related to poor outcome in patients with anal fistula (AF). Due to being associated with short-term and long-term cure rate, the perioperative complications have been receiving widespread attention following AF surgery. This study aims to identify a set of predictive factors to develop risk prediction models for recurrence and related complications following AF surgery. We plan to develop and validate risk prediction models, using information collected through a WeChat patient-reported questionnaire system combined with clinical, laboratory and imaging findings from the perioperative period until 3-6 months following AF surgery.

Methods and analysis

This is a prospective hospital-based cohort study using a linked database collected health data as well as the follow-up outcomes for all adult patients who suffered from AF at a tertiary referral hospital in Shanghai, China. We will perform logistic regression models to predict anal fistula recurrence (AFR) as well as related complications (eg, wound hemorrhage, fecal impaction, urinary retention, delayed wound healing and unplanned hospitalization) during and after AF surgery, and machine learning approaches will also be applied to develop risk prediction models. This prospective study aims to develop the first risk prediction models for AFR and related complications using multi-dimensional variables. These tools can be used to

For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

warn, motivate and empower patients to avoid some modifiable risk factors to early prevent postoperative complications. This study will also provide alternative tools for early screening of high-risk patients of AFR and related complications, helping surgeons better understand the aetiology and outcome of AF in an earlier stage.

Ethics and dissemination

The study is approved by the Institutional Review Board of Shuguang Hospital affiliated with Shanghai University of Traditional Chinese Medicine (Approval Number: 2019-699-54-01). The results of this study will be submitted to international scientific peer-reviewed journals or conferences in surgery, anorectal surgery or anorectal diseases.

Trial registration number

Chinese Clinical Trial Registry (ChiCTR1900025069); Pre-results.

Key words

Anal fistula; treatment outcome; recurrence; surgery; cohort study; prediction model

For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

Strengths and limitations of this study

- This is the first large prospective cohort study of patients with anal fistula at a tertiary referral hospital in China.
- A higher events per candidate predictive variable (⩾20) will be applied which can generally eliminate bias in regression coefficients for prediction models and guarantee a sufficient sample size for model development.
- Candidate predictors will be identified from the published and updated systematic reviews, expert opinions from Delphi surveys, univariable or multivariable logistic regression analysis.
- Bootstrapping procedure will be applied for the internal and external validation of the prediction models.
- A higher probability of missing data due to non-response bias may occur as many of the variables are collected through a WeChat questioneer system.

For peer review only

For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

Introduction

Anal fistula (AF) is common perianal condition defined by a pathological epithelial tract that connects the anal canal or rectum and the surface of the perianal region, which is also regarded as a chronic stage of perianal abscess.[1] Postoperative recurrence, defined as persistence or recurrence of AF symptoms, or the development of recurrent perianal sepsis or chronic AF within six months of surgery,[2 3] is one of the consequences which can be related to a bad surgical procedure but may also be due to the insidiousness of the disease. Our recently published meta-analysis based on 20 studies reported a recurrence rate of about 19% (95% CI 0.15-0.23) in patients having AF surgery.[4] Because of the high degree of difficulty of surgery for patients with high complex AF, the postoperative recurrence rate of these patients can be as high as 50%, and the failure rate of reoperation remains 10%.[5-8] It is considered as one of the most difficult and complicated anorectal diseases.

A large number of studies have shown that the recurrence of AF is related to multiple factors, such as unclear diagnosis or failure to dealing with the correct internal orifice, blind stump of fistula, incorrect method of seton, omission of branch of the fistula and poor drainage.[9 10] Studies also reported that anal fistula recurrence (AFR) was associated with individual patient characteristics, such as history of the enteritis, previous anal surgery, obesity and smoking.[4 11-13] Li et.al retrospectively analyzed 1783 patients with AF receiving surgical treatment and found that the location of AF, previous perianal surgery, seton history and enteritis were

For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

independent risk factors for AFR.[5] Recently, according to the evidence grading criteria based on Egger's P value, total sample size and between-study heterogeneity, we published a meta-analysis involving 20 studies with 6168 patients and concluded high transsphincteric fistula, internal opening unidentified, and horseshoe extensions were independent risk factors for AFR with high-quality evidence, while prior anal surgery, seton placement surgery, and multiple fistula tract were demonstrated as risk factors for AFR with moderate-quality evidence.[4]

Factors influencing other perioperative complications related AF surgery including wound hemorrhage, fecal impaction, urinary retention, delayed wound healing and unplanned hospitalization are also rarely reported. Therefore, there is an urgent need to develop risk prediction tools for the complete profile of risk factors for AFR and those related complications.

Which AF patients will be cured after surgery and which ones will not, are rarely investigated. The development of a prediction model for AFR following surgery to identify those patients with a higher risk of developing complications during follow up, would be of significant importance. Firstly, surgeons can provide patients preoperatively about an estimated surgical cure rate according to the prediction models. Moreover, the current knowledge in the literature reporting potential predictive factors could help patients know well their individual risk factors

For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

and avoid modifiable ones in order to improve the cure rate, which have been well described and applied in the prevention of other diseases. [14-17]

However, so far, there are no effective screening tools to evaluate and predict the risk of recurrence or other adverse outcomes of AF. Therefore, the aim of the current study was to develop and validate multivariable prediction models that predict postoperative AFR and related complications.

Aims and objectives

The aim of this study is to develop risk prediction models for postoperative recurrence as well as other surgery-related complications in a prospective hospital-based AF cohort. Risk prediction model for perioperative complications will also be developed. Flowchart of prediction model development and assessment is provided in Figure 1.

The detailed tasks of this study are to:

1. Calculate the 3 to 6-month incidence of recurrence, and related complications in patients following AF surgery.

2. Establish the risk factors that significantly predict postoperative AFR and related complications based on the AF cohort in a tertiary referral center.

3. Develop and validate the risk prediction models for postoperative AFR and related complications.

For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

4. Considering the different scenario for different surgical interventions, stratified analyses are conducted based on surgery type. If possible, risk prediction model will also be developed in relevant sub-populations, such as those only receiving fistulectomy or fistulotomy, which can account for more than 60% of our AF cohort populations.

We also have the following two hypotheses examined:

1. Patient-related demographic characteristics, fistula and surgery-related factors are predictive of postoperative AFR and related complications as dependent variables.

2. The risk prediction models for postoperative AFR and related complications developed in our study have more than 70% of discriminating power.

Patients and Methods

Study design and participants

This study is a single-center, prospective observational study on a hospital-based cohort enrolled at a tertiary referral center in Shanghai, China.

Eligibility criteria

The enrollment of the cohort subject was initiated from June, 2019. All subjects who will undergo surgical intervention for AF will be included for inclusion. All operations will be performed by a group of colon and rectal surgeons at Shuguang Hospital, a regional tertiary referral center. Exclusion criteria are those whose age <

For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

18 years, non-cryptoglandular fistula (eg, anal fistula due to inflammatory bowel disease, human immunodeficiency virus, malignant cancer, or obstetrical trauma), and rectovaginal or rectourethral fistula. The electronic medical records of the included subjects should be complete.

Trained clinical investigators are collecting data in several categories, including baseline demographics, laboratory examinations, clinical data, imaging findings and follow-up information 3 to 6 months postoperatively. Planned clinical reviews or electronic surveys are conducted during hospitalization and every half to 3 months after discharge for 6 months.

Data collection

The research team comprised a principal investigator, 5 to 8 anorectal surgeons who are trained and supervised by the Ethics committee of Shuguang Hospital. Written informed consent was obtained from all patients. Investigators will not intervene in any aspects of patient surveys at every stage of data collection and follow-up. Data are collected using a convenient follow-up system supported by Empower EDC (Solutions, Boston, Massachussetts, USA). This electronic system introduces a machine learning algorithm, through which we can use the data already entered in the Empower system to train the algorithm model and let the system itself develop quality control algorithms, validate the entered data and identify missing or suspicious data. Finally, the data manager will check the missing or suspicious data, confirm their

For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

completeness and asked the data manager to provide additional data when necessary. Furthermore, an automatic reminder follow-up function also plays a pivotal role during the whole follow-up period.

Patient and public involvement

Patients and public will not be involved in the development, design, conduct or reporting of the study. The general results will be disseminated to participants through public education during follow-up.

Clinical outcomes

The primary study end point is postoperative recurrence following AF surgery defined as persistence or recurrence of AF symptoms, or the development of recurrent perianal sepsis or chronic AF within 3 to 6 months of surgery.[2 3 18] The second end point is a composite outcome of postoperative comorbidities or any of the following equivalent events including AFR, wound hemorrhage, fecal impaction, urinary retention, delayed wound healing or unplanned hospitalization associated with AF surgery. Outcomes were ascertained by the treating clinician combined with outpatient medical records or patient self-reports.

Selection of predictor variables

Candidate variables for the prediction model of the composite outcome of postoperative comorbidities will be screened according to the following pre-set

For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

criteria: (1) prior clinical knowledge; (2) results from a systematic review updated in November 2019 based on our published one[4] with sufficient evidence to include them as predictive variables in the risk model for AFR as is demonstrated below; or (3) agreed upon by a group of anorectal surgeons or experts for their clinical relevance using a two-round Delphi survey. We initially identified the following covariates as relevant candidate variables based on the systematic reviews as well as clinical knowledge and/or relevance. The determination of all other candidate variables are based on results of post-hoc analysis using univariable or multivariable survival analyses with a threshold of $p < 0.05$.

Factors identified from the systematic reviews and Delphi survey (manuscript under review), will be measured at baseline. These include factors involving the identified significant risk factors which are reported in our meta-analysis are presented as follows:

- ► Prior anal surgery.
- ► Seton placement surgery.
- ► High transsphincteric fistula.
- ► Internal opening unidentified.
- ► Horseshoe extensions.
- ► Multiple fistula tracts.

For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

Some of the demographic factors and surgery details will also be collected due to limited power in the literature reviews and some non-significant potential factors (eg, smoking or alcohol use) may be risk factors and are also included as follows:

► Gender.

► Age.

► Smoking.

► Alcohol use.

► Diabetes mellitus.

► Obesity.

► Preoperative seton drainage.

►High internal opening.

►Postoperative drainage.

►Supralevator extensions.

Other factors like laboratory examinations and MR imaging parameters (height of the internal openings, height and number of fistula, etc.) will also be collected. Moreover, some other factors like chronic steroid therapy, diverting stoma, the surgeon's level of training, postoperative bowel confinement and antibiotic prophylaxis reported in previous literature are selected for regression analysis as well. In addition, relevant factors from the expert-opinion survey were also assessed including the number of prior anal fistula surgeries, types of surgery performed such as staged fistulotomies, endorectal advancement flap and ligation of the intersphinteric fistula tract with and

For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

without seton drains, some nutrition parameters and immunomodulation medication use.

Categorization of potential predictors

For categorical predictors, we can code them as "factor" variables, with coding as dummy variables, for example, smoking is coded originally as "1" for never smoker, "2" for past smoker, and "3" for current smoker and never smoker was selected as the reference category. Similar manner can be applied with alcohol use.[19] Continuous variables formally should be measured at an interval or ratio scale, and should be able to take any value in a range. We treat ordered variables as linear which is generally reasonable for prediction. In other cases, continuous predictors can be grouped with meaningful categorization; for example, body mass index (BMI) can be classified based on internationally recognized categories (i.e., underweight, normal weight, overweight, and obesity).[20] Based on previous experiences, we will be deriving some predictors based on the responses of the surveys. However, in case some subjectivity in the classifications of these predictors may occur, sensitivity analyses will be performed to examine the robustness of our definitions during model development and validation.

Study quality control for the prediction models

Based on the summary of methodological quality and developmental stage of prediction models by van Oort et al., we are developing predesigned criteria for

For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

quality control of our prediction models, which can make us carry out the study and report the results more rigorously.[21 22] The methodological checklist of the study are presented in supplementary material.

Missing data

Candidate predictors with more than 60% missingness will be excluded. For those less than 60% missingness, multiple imputation are to be performed by imputing 20 complete data sets using multivariate normal regression, [23-26] which can reasonably approximate the true distributional relationship between the missing values and the available ones.[27] Among various multiple imputation approaches, fully conditional specification (FCS) and multivariate normal imputation (MVNI) are preferred, because they have been proved to be generally less biased than complete-case analysis. They can both generate similar results in the presence of either binary or ordinal variables that are not generally normally distributed. [27]

Statistical analysis for model derivation

Logistic regression will be applied to develop our prediction models for the binary outcomes. All data processing and statistical analysis will be performed using EmpowerStats software (www. empowerstats.com; X&Y Solutions, Inc., Boston, MA, USA) and statistical software packages R (R Foundation, Vienna, Austria).

We will first study the association between each potential variable and the outcome based on univariable analysis. Variables are considered further for multivariable

For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

regression modelling when they are associated with a p-value less than 0.20. Normality or linearity will be evaluated for the continuous predictors. Fractional polynomials are advocated for associations between the continuous predictors and the outcome for non-linear relationships. [28 29] We will perform backward stepwise selection with a $p < 0.001$ as the inclusion threshold and a $p > 0.05$ as the exclusion threshold for each imputed data set.

Predictors which appear in the imputation models with an inclusion fraction of ⩾50% are qualified for the final multivariable model. Though there is no consensus regarding the optimal method for selecting predictors for inclusion, backwards elimination is generally considered as the preferred procedure as reported by Mantel et al.[30] Forward stepwise procedure will also be performed to repeat the analysis to test the robustness of the models. Overall regression coefficient estimates of the models will be generated with the combination of the imputed datasets based on Rubin's Rules, while taking into account uncertainty in the imputed values.[23 24 26] Collinearity will also be assessed which refers to the fact that predictors can have strong correlation with each other, defined as correlation coefficient >0.8, or variance inflation factor >10.[31] Then we will examine the interactions among the regression models.

Prediction model performance assessment

For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

Prediction models will be developed with a random sample of 60% of the AF cohort as the derivation cohort, and then validated with the remaining sample of 40% of the cohort as the validation cohort. The predictive performance in the derivation and validation cohort will be evaluated and reported by examining measures of predictive accuracy, discrimination and calibration. Nagelkerke's R^2 and Brier score will be used for the measurement of predictive accuracy.[32 33] The discriminative ability of the prediction models are evaluated using several statistics, which are according to the discriminative and calibration ability in both derivation and validation AF cohort. Model discrimination means the ability of the models to differentiate between high-risk patients and low-risk patients (having high or low risk of AFR or surgery-related complications). This will be assessed via Harrell's concordance statistic (C-index).[34] The calculation of the C-index will be performed in each of the 20 imputed data sets, and then averaged based on Rubin's rule.[35] The model is interpreted as having no discriminatory ability when a value of C-index is 0.5, and has perfect discrimination when a value of C-index is 1.0.[34] Calibration implies the agreement between the predicted outcomes and the observed outcomes, which is evaluated with the Hosmer and Lemeshow test for goodness of fit in all imputed datasets presented with calibration plots.[36] Calibration-in-the-large, which defines as the agreement between mean observed outcomes and mean predictions, will also be assessed for calibration.[37]

Internal and external validation of prediction model

For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

To make the prediction models reproducible, we have to conduct internal validation. Bootstrapping technique, as one of the most attractive resampling techniques, is a mostly applied validation method, which seems to be most efficient for obtaining stable optimism-corrected estimates.[34 38] It has been reported that bootstrap validation is a feasible technique for most prediction models with at least a 500 bootstrap resampling procedure using Harrell's validate function, which can adjust the developed models for over-fitting.[39] We will also apply temporal validation as external validation using a more recent AF patient cohort. [40]

Sample size

Since there are no widely accepted methods for the estimation of the sample size requirements to develop the risk prediction models, the size of this AF cohort will be calculated to have 20 events per candidate predictive variable (EPV, defined as the ratio of the number of individuals with the outcome event to the number of candidate predictors), which can generally eliminate bias in regression coefficients for prediction models with low-prevalence binary predictor development (the estimated recurrence rate <20%) and adequately power the logistic regression models.[41,42] According the findings by Ogundimu et al. [43], a higher EPV (⩾20) can generally eliminate bias in regression coefficients for prediction models with low-prevalence binary predictors development. Then we estimate 400 events allow for 20 predictor variables (EPV=20). Considering 5-20% recurrence rate, we assume that at least 4000-8000 patients should be collected for model development. In addition, surgery

For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

type and fistula type-stratified analyses will also be performed to examine the different effect of these factors on disease recurrence or other related complications in each subgroup. Risk prediction models, if possible, can also be developed in those sub-populations. The cohort size with more than 4000-8000 patients will provide sufficient power to perform those analyses and develop prediction models in those subgroups.

Follow-up and methodological quality control

The application of WeChat questionnaires to collect data will inevitably increase the probability of missing data. However, we have made some pre-designed countermeasures. For example, we have set up follow-up reminders via the WeChat questionnaire system. Moreover, every week two trained clinical fellows cross-check the data, and will contact the respondents by phone or WeChat about the missing contents, which can minimize the missing data and lost to follow-up rate.

Limited studies have identified specific criteria for quality control in a prediction model, but we have strictly adhered to the guidelines for the reporting of studies developing, validating multivariable clinical prediction models as is reported in the TRIPOD (Transparent Reporting of a multivariable prediction model for Individual Prognosis Or Diagnosis) Statement to ensure methodological rigour.[44] All issues have been addressed in this study in supplementary material.

For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

Ethics and dissemination

The study is approved by the Institutional Review Board of Shuguang Hospital affiliated with Shanghai University of Traditional Chinese Medicine (Approval Number: 2019-699-54-01). The results of this cohort will be submitted to international scientific peer-reviewed journals or conferences in surgery, anorectal surgery or anorectal diseases.

Discussion

In this study, we plan to develop internally validated models for the prediction of recurrrence as well as postoperative complications among AF patients. The models will be developed based on a large AF cohort in a hospital-representative linked database with validated clinical information. The collected variables include WeChat questionnaires, clinical, laboratory and imaging findings, as well as follow-up information which are routinely being gathered at the time of enrollment.

According to the existing knowledge and systematic reviews, it is highly plausible that a number of patient, fistula and surgery related characteristics (eg, patient-related variables such as gender, age, diabetes mellitus or obesity,[45-47] lifestyle factors such as smoking, alcohol abuse,[48 49] fistula-related factors such as number of fistula tracts, horseshoe extensions, classification and location of fistula, surgery-related variables such as prior anal surgery and postoperative drainage[48 50 51]) easily ascertainable before surgery may predict AFR. Similar risk prediction models exist in other

For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

diseases, such as the Framingham Risk Score model to predict cardiovascular disease risk[52] and the Korean Crohn's Disease Prediction (KCDP) model to predict the clinical course of Crohn's disease.[53] Until now, risk factor investigation of predictors of perioperative surgery-related complications has been limited to assessment of single predictors with small sample size. The risk prediction models can help inform surgeons regarding high risk AF patients based on the overall risk factors. The primary purpose of study was to develop two risk prediction models to facilitate surgeons in identifying AF patients who will have surgical treatment at higher risk of developing recurrence and surgery-related complications. The predictive models will help both clinicians and patients identify the risk of complications after AF surgery in advance, take necessary interventions to reduce the risk of surgery-related complications as well as the personal and social financial burden brough about by those complications. The accurate risk prediction models are especially instructive for the development of the optimal surgical plan to achieve optimal surgical outcome.

Our study has several strengths. To the best of our knowledge, this is the first study with the primary aim to develop, internally and externally validate multivariable prediction models for AFR and related complications following AF surgery. Multidimensional clinically useful candidate predictors will be fully examined from a variety of sources including our published and updated systematic reviews, expert opinions from Delphi surveys and univariable or multivariable logistic regression analysis. Second, we will apply the internal and external validation of the prediction

For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

models using bootstrapping procedure. Third, multiple imputation will also be used to treat the missing data. Last but not least, our study is a prospective cohort one with adequate follow-up period which can minimize certain bias.

Our study also has limitations. As many of the variables are collected through a WeChat questioneer system during the hospitalization and follow-up, a higher probability of missing data due to non-response bias may occur. To solve this issue, we reguarly send reminders to those who does not respond after discharge. Second, though we will investigate a series of potential predictors, some more potentially predictors will not be involved or not collected in the current study, data related to postoperative nursing stratergy and outpatient follow-up frequency for example. Moreover, bias may also result from the single-center recruitment of our study and will be improved through multicenter recruitment in the future. Last but not least, one key potential issue that needs to be considered a priori is the variable recurrence rate depending on the risk factors included and identified in the testing and validating cohorts since this may affect the C-index. It is possible that a lack of key variable inclusions in the models may result in decreased discriminatory ability. This is a function of the database and points of interest included that may need to be maximized before proceeding with development of the testing model.

The newly developed risk algorithms may have significant applications in clinical practice by helping recommend optimal surgical approach for a specific AF patient,

For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

as well as intensive perioperative care and education, timely assessment and discussion of the need for interventions to those most at high risk of developing recurrence or surgery-related complications. The models will specifically identify AF patients who are likely to develop recurrence or related complications following AF surgery to offer the quantitative evaluation of the risk. Moreover, the models will also provide reference information for preventing recurrence and reducing the rate of recurrence after operation, and to intervene some high risk factors in the early stage.

In summary, this study protocol summarizes the design of development and validation studies for a risk screening tool in patients receiving AF surgery. Results from this study will be interpreted for the purpose of clinical decision making. The models to be developed of the study could be used to make new recommendations for perioperative AF patients.

For peer review only

For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

References

1. Tantiphlachiva K. 7 - Anorectal Abscess and Fistula. In: Coss-Adame E, Remes-Troche JM, eds. Anorectal Disorders: Academic Press 2019:81-100.

2. Abbas MA, Jackson CH, Haigh PI. Predictors of outcome for anal fistula surgery. Archives of surgery (Chicago, Ill : 1920) 2011;146(9):1011-6. doi: 10.1001/archsurg.2011.197

3. Hamadani A, Haigh P, Liu I, et al. Who is at risk for developing chronic anal fistula or recurrent anal sepsis after initial perianal abscess? Dis Colon Rectum 2009;52(2):217-21.

4. Mei Z, Wang Q, Zhang Y, et al. Risk Factors for Recurrence after anal fistula surgery: A meta-analysis. International Journal of Surgery 2019;69:153-64. doi: https://doi.org/10.1016/j.ijsu.2019.08.003

5. Li J, Yang W, Huang Z, et al. [Clinical characteristics and risk factors for recurrence of anal fistula patients]. Zhonghua wei chang wai ke za zhi = Chinese journal of gastrointestinal surgery 2016;19(12):1370-74.

6. Gottgens KW, Smeets RR, Stassen LP, et al. Systematic review and meta-analysis of surgical interventions for high cryptoglandular perianal fistula. International journal of colorectal disease 2015;30(5):583-93. doi: 10.1007/s00384-014-2091-8

For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

7. Gottgens KW, Janssen PT, Heemskerk J, et al. Long-term outcome of low perianal fistulas treated by fistulotomy: a multicenter study. International journal of colorectal disease 2015;30(2):213-9. doi: 10.1007/s00384-014-2072-y

8. Bakhtawar N, Usman M. Factors Increasing the Risk of Recurrence in Fistula-in-ano. Cureus 2019;11(3):e4200. doi: 10.7759/cureus.4200

9. Sygut A, Mik M, Trzcinski R, et al. How the location of the internal opening of anal fistulas affect the treatment results of primary transsphincteric fistulas. Langenbeck's archives of surgery 2010;395(8):1055-9. doi: 10.1007/s00423-009-0562-0

10. Dziki A, Bartos M. Seton treatment of anal fistula: experience with a new modification. The European journal of surgery = Acta chirurgica 1998;164(7):543-8. doi: 10.1080/110241598750005930

11. Hermann J, Eder P, Banasiewicz T, et al. Current management of anal fistulas in Crohn's disease. Przeglad gastroenterologiczny 2015;10(2):83-8. doi: 10.5114/pg.2015.49684

12. Hamadani A, Haigh PI, Liu IL, et al. Who is at risk for developing chronic anal fistula or recurrent anal sepsis after initial perianal abscess? Diseases of the colon and rectum 2009;52(2):217-21. doi: 10.1007/DCR.0b013e31819a5c52

For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

13. Ellis CN, Clark S. Effect of tobacco smoking on advancement flap repair of complex anal fistulas. Diseases of the colon and rectum 2007;50(4):459-63. doi: 10.1007/s10350-006-0829-2

14. Moons KG, Altman DG, Vergouwe Y, et al. Prognosis and prognostic research: application and impact of prognostic models in clinical practice. Bmj 2009;338:b606. doi: 10.1136/bmj.b606

15. Ming C, Viassolo V, Probst-Hensch N, et al. Machine learning techniques for personalized breast cancer risk prediction: comparison with the BCRAT and BOADICEA models. Breast cancer research : BCR 2019;21(1):75. doi: 10.1186/s13058-019-1158-4

16. Khan SS, Ning H, Shah SJ, et al. 10-Year Risk Equations for Incident Heart Failure in the General Population. Journal of the American College of Cardiology 2019;73(19):2388-97. doi: 10.1016/j.jacc.2019.02.057

17. Terry MB, Liao Y, Whittemore AS, et al. 10-year performance of four models of breast cancer risk: a validation study. The Lancet Oncology 2019;20(4):504-17. doi: 10.1016/S1470-2045(18)30902-1

18. Yano T, Asano M, Matsuda Y, et al. Prognostic factors for recurrence following the initial drainage of an anorectal abscess. International journal of colorectal disease 2010;25(12):1495-8. doi: 10.1007/s00384-010-1011-9 [published Online First: 2010/07/20]

For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

19. Steyerberg EW. Coding of Categorical and Continuous Predictors. In: Steyerberg EW, ed. Clinical Prediction Models: A Practical Approach to Development, Validation, and Updating. Cham: Springer International Publishing 2019:175-90.

20. Morgan TM, Elashoff RM. Effect of Categorizing a Continuous Covariate on the Comparison of Survival Time. Journal of the American Statistical Association 1986;81(396):917-21. doi: 10.1080/01621459.1986.10478350

21. van Oort L, van den Berg T, Koes BW, et al. Preliminary state of development of prediction models for primary care physical therapy: a systematic review. Journal of Clinical Epidemiology 2012;65(12):1257-66. doi: https://doi.org/10.1016/j.jclinepi.2012.05.007

22. Rushton AB, Verra ML, Emms A, et al. Development and validation of two clinical prediction models to inform clinical decision-making for lumbar spinal fusion surgery for degenerative disorders and rehabilitation following surgery: protocol for a prospective observational study. BMJ Open 2018;8(5):e021078. doi: 10.1136/bmjopen-2017-021078

23. Wood AM, White IR, Royston P. How should variable selection be performed with multiply imputed data? Statistics in Medicine 2008;27(17):3227-46. doi: 10.1002/sim.3177

For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

24. White IR, Royston P, Wood AM. Multiple imputation using chained equations: Issues and guidance for practice. Statistics in Medicine 2011;30(4):377-99. doi: 10.1002/sim.4067

25. Schafer JL. Multiple imputation: a primer. Statistical Methods in Medical Research 1999;8(1):3-15. doi: 10.1177/096228029900800102

26. Marshall A, Altman DG, Holder RL, et al. Combining estimates of interest in prognostic modelling studies after multiple imputation: current practice and guidelines. BMC Medical Research Methodology 2009;9(1):57. doi: 10.1186/1471-2288-9-57

27. Lee KJ, Carlin JB. Multiple Imputation for Missing Data: Fully Conditional Specification Versus Multivariate Normal Imputation. American Journal of Epidemiology 2010;171(5):624-32. doi: 10.1093/aje/kwp425

28. Sauerbrei W, Royston P, Binder H. Selection of important variables and determination of functional form for continuous predictors in multivariable model building. Statistics in Medicine 2007;26(30):5512-28. doi: 10.1002/sim.3148

29. Royston P, Sauerbrei W. Building multivariable regression models with continuous covariates in clinical epidemiology--with an emphasis on fractional polynomials. Methods of information in medicine 2005;44(4):561-71.

30. Mantel N. Why Stepdown Procedures in Variable Selection. Technometrics 1970;12(3):621-25. doi: 10.1080/00401706.1970.10488701

For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

31. Vittinghoff E, Glidden DV, Shiboski SC, et al. Regression methods in biostatistics: linear, logistic, survival, and repeated measures models: Springer Science & Business Media 2011.

32. Steyerberg EW, Vickers AJ, Cook NR, et al. Assessing the performance of prediction models: a framework for traditional and novel measures. Epidemiology 2010;21(1):128-38. doi: 10.1097/EDE.0b013e3181c30fb2

33. NAGELKERKE NJD. A note on a general definition of the coefficient of determination. Biometrika 1991;78(3):691-92. doi: 10.1093/biomet/78.3.691

34. Harrell Jr FE. Regression modeling strategies: with applications to linear models, logistic and ordinal regression, and survival analysis: Springer 2015.

35. Rubin DB. Multiple imputation for nonresponse in surveys: John Wiley & Sons 2004.

36. Hosmer Jr DW, Lemeshow S, Sturdivant RX. Applied logistic regression: John Wiley & Sons 2013.

37. Snell KI, Ensor J, Debray TP, et al. Meta-analysis of prediction model performance across multiple studies: Which scale helps ensure between-study normality for the C-statistic and calibration measures? Statistical methods in medical research 2018;27(11):3505-22.

For peer review only

38. Chatfield C. Model uncertainty, data mining and statistical inference. Journal of the Royal Statistical Society: Series A (Statistics in Society) 1995;158(3):419-44.

39. Steyerberg EW, Bleeker SE, Moll HA, et al. Internal and external validation of predictive models: a simulation study of bias and precision in small samples. Journal of clinical epidemiology 2003;56(5):441-47.

40. Justice AC, Covinsky KE, Berlin JA. Assessing the generalizability of prognostic information. Annals of internal medicine 1999;130(6):515-24.

41. Concato J, Peduzzi P, Holford TR, et al. Importance of events per independent variable in proportional hazards analysis I. Background, goals, and general strategy. Journal of clinical epidemiology 1995;48(12):1495-501.

42. Steyerberg EW. Clinical Prediction Models: A Practical Approach to Development, Validation, and Updating. SPRINGER NATURE, 2019.

43. Ogundimu EO, Altman DG, Collins GS. Adequate sample size for developing prediction models is not simply related to events per variable. J Clin Epidemiol. 2016;76:175–182.

44. Moons KG, Altman DG, Reitsma JB, et al. Transparent Reporting of a multivariable prediction model for Individual Prognosis or Diagnosis (TRIPOD): explanation and elaboration. Annals of internal medicine 2015;162(1):W1-W73.

For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

45. Hamadani A, Haigh PI, Liu I-LA, et al. Who is at risk for developing chronic anal fistula or recurrent anal sepsis after initial perianal abscess? Diseases of the colon & rectum 2009;52(2):217-21.

46. Sonoda T, Hull T, Piedmonte MR, et al. Outcomes of primary repair of anorectal and rectovaginal fistulas using the endorectal advancement flap. Diseases of the colon & rectum 2002;45(12):1622-28.

47. Hyman N, O'Brien S, Osler T. Outcomes after fistulotomy: results of a prospective, multicenter regional study. Diseases of the Colon & Rectum 2009;52(12):2022-27.

48. Ellis CN, Clark S. Effect of tobacco smoking on advancement flap repair of complex anal fistulas. Diseases of the colon & rectum 2007;50(4):459-63.

49. Schwandner T, Roblick MH, Kierer W, et al. Surgical treatment of complex anal fistulas with the anal fistula plug: a prospective, multicenter study. Diseases of the Colon & Rectum 2009;52(9):1578-83.

50. Garcia-Aguilar J, Belmonte C, Wong WD, et al. Anal fistula surgery: Factors associated with recurrence and incontinence. Diseases of the colon & rectum 1996;39(7):723-29.

51. Athanasiadis S, Köhler A, Nafe M. Treatment of high anal fistulae by primary occlusion of the internal ostium, drainage of the intersphincteric space, and mucosal advancement flap. International journal of colorectal disease 1994;9(3):153-57.

For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

52. Fox ER, Samdarshi TE, Musani SK, et al. Development and validation of risk prediction models for cardiovascular events in black adults: the Jackson Heart Study Cohort. JAMA cardiology 2016;1(1):15-25.

53. Park Y, Cheon JH, Park YL, et al. Development of a novel predictive model for the clinical course of Crohn's disease: results from the connect study. Inflammatory bowel diseases 2017;23(7):1071-79.

For peer review only

For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

Figure legend

Figure 1. Flowchart of prediction model development and assessment.

For peer review only

For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

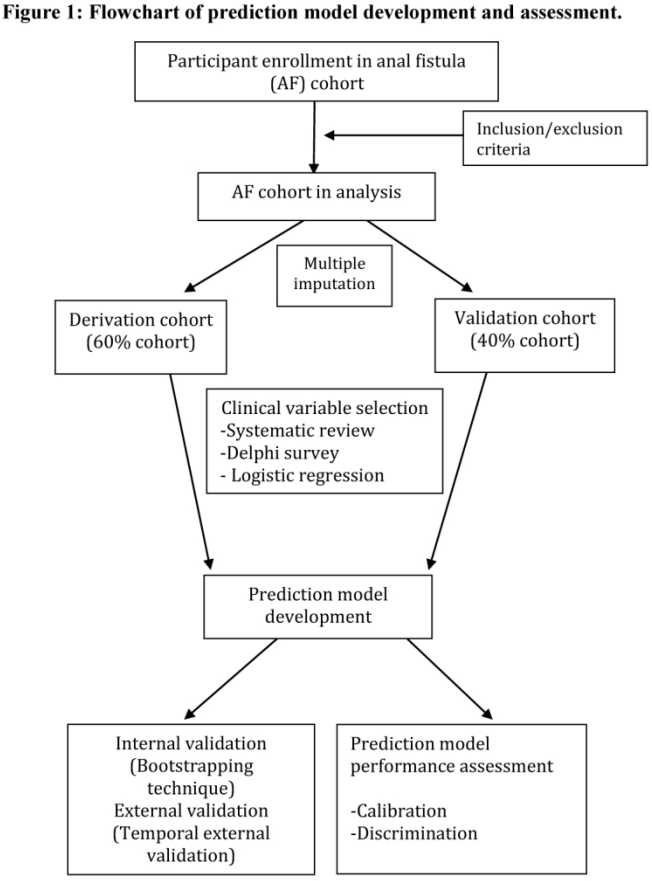

Figure 1. Flowchart of prediction model development and assessment.

BMJ Open: first published as 10.1136/bmjopen-2019-035134 on 4 March 2020. Downloaded from http://bmjopen.bmj.com/ on April 20, 2024 by guest. Protected by copyright.

TRIPOD Checklist: Prediction Model Development and Validation

This document contains a completed TRIPOD checklist for the manuscript: "**Development of screening tools to predict the risk of recurrence and related complications following anal fistula surgery: protocol for a prospective cohort study"**.

Zubing Mei, Yue Li, Zhijun Zhang, Haikun Zhou, Suzhi Liu, Ye Han, Peixin Du, Xiufang Qin, Zhuo Shao, Maojun Ge, Qingming Wang, Wei Yang.

As this TRIPOD checklist refers to a study protocol, not all items are relevant at this stage. We have tried to the furthest extent possible to make the protocol adhere to the TRIPOD checklist and all reporting of results will be in accordance to the protocol and the TRIPOD statement. Page numbers in the submitted manuscript are provided. For items that are only partly relevant at this time, page numbers are provided in parentheses and for items that are not relevant at this time a "-" has been written.

Section/Topic	Item		Checklist Item	Page
Title and abstract				
Title	1	D;V	Identify the study as developing and/or validating a multivariable prediction model, the target population, and the outcome to be predicted.	1
Abstract	2	D;V	Provide a summary of objectives, study design, setting, participants, sample size, predictors, outcome, statistical analysis, results, and conclusions.	5-6
Introduction				
Background and objectives	3a	D;V	Explain the medical context (including whether diagnostic or prognostic) and rationale for developing or validating the multivariable prediction model, including references to existing models.	8-9
	3b	D;V	Specify the objectives, including whether the study describes the development or validation of the model or both.	8-10
Methods				
Source of data	4a	D;V	Describe the study design or source of data (e.g., randomized trial, cohort, or registry data), separately for the development and validation data sets, if applicable.	11
	4b	D;V	Specify the key study dates, including start of accrual; end of accrual; and, if applicable, end of follow-up.	11-12
Participants	5a	D;V	Specify key elements of the study setting (e.g., primary care, secondary care, general population) including number and location of centres.	12-13
	5b	D;V	Describe eligibility criteria for participants.	11-12
	5c	D;V	Give details of treatments received, if relevant.	11
Outcome	6a	D;V	Clearly define the outcome that is predicted by the prediction model, including how and when assessed.	13
	6b	D;V	Report any actions to blind assessment of the outcome to be predicted.	13
Predictors	7a	D;V	Clearly define all predictors used in developing or validating the multivariable prediction model, including how and when they were measured.	13-15
	7b	D;V	Report any actions to blind assessment of predictors for the outcome and other predictors.	15
Sample size	8	D;V	Explain how the study size was arrived at.	20
Missing data	9	D;V	Describe how missing data were handled (e.g., complete-case analysis, single imputation, multiple imputation) with details of any imputation method.	17
Statistical analysis methods	10a	D	Describe how predictors were handled in the analyses.	17-18
	10b	D	Specify type of model, all model-building procedures (including any predictor selection), and method for internal validation.	17-20
	10c	V	For validation, describe how the predictions were calculated.	20
	10d	D;V	Specify all measures used to assess model performance and, if relevant, to compare multiple models.	19
	10e	V	Describe any model updating (e.g., recalibration) arising from the validation, if done.	-
Risk groups	11	D;V	Provide details on how risk groups were created, if done.	-
Development vs. validation	12	V	For validation, identify any differences from the development data in setting, eligibility criteria, outcome, and predictors.	19-20
Results				
Participants	13a	D;V	Describe the flow of participants through the study, including the number of participants with and without the outcome and, if applicable, a summary of the follow-up time. A diagram may be helpful.	Figure 1
	13b	D;V	Describe the characteristics of the participants (basic demographics, clinical features, available predictors), including the number of participants with missing data for predictors and outcome.	-
	13c	V	For validation, show a comparison with the development data of the distribution of important variables (demographics, predictors and outcome).	-
Model development	14a	D	Specify the number of participants and outcome events in each analysis.	-
	14b	D	If done, report the unadjusted association between each candidate predictor and outcome.	-
Model	15a	D	Present the full prediction model to allow predictions for individuals (i.e., all regression	-

For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

TRIPOD Checklist: Prediction Model Development and Validation

specification			coefficients, and model intercept or baseline survival at a given time point).	-
	15b	D	Explain how to the use the prediction model.	-
Model performance	16	D;V	Report performance measures (with CIs) for the prediction model.	-
Model-updating	17	V	If done, report the results from any model updating (i.e., model specification, model performance).	-
Discussion				
Limitations	18	D;V	Discuss any limitations of the study (such as nonrepresentative sample, few events per predictor, missing data).	24
Interpretation	19a	V	For validation, discuss the results with reference to performance in the development data, and any other validation data.	-
	19b	D;V	Give an overall interpretation of the results, considering objectives, limitations, results from similar studies, and other relevant evidence.	-
Implications	20	D;V	Discuss the potential clinical use of the model and implications for future research.	24-25
Other information				
Supplementary information	21	D;V	Provide information about the availability of supplementary resources, such as study protocol, Web calculator, and data sets.	-
Funding	22	D;V	Give the source of funding and the role of the funders for the present study.	2

*Items relevant only to the development of a prediction model are denoted by D, items relating solely to a validation of a prediction model are denoted by V, and items relating to both are denoted D;V. We recommend using the TRIPOD Checklist in conjunction with the TRIPOD Explanation and Elaboration document.

For peer review only

TRIPOD Checklist: Prediction Model Development and Validation

This document contains a completed TRIPOD checklist for the manuscript: "**Development of screening tools to predict the risk of recurrence and related complications following anal fistula surgery: a prospective cohort study"**.

Zubing Mei, Yue Li, Zhijun Zhang, Suzhi Liu, Haikun Zhou, Ye Han, Peixin Du, Zhuo Shao, Maojun Ge, Qingming Wang, Wei Yang.

As this TRIPOD checklist refers to a study protocol, not all items are relevant at this stage. We have tried to the furthest extent possible to make the protocol adhere to the TRIPOD checklist and all reporting of results will be in accordance to the protocol and the TRIPOD statement. Page numbers in the submitted manuscript are provided. For items that are only partly relevant at this time, page numbers are provided in parentheses and for items that are not relevant at this time a "-" has been written.

Section/Topic	Item		Checklist Item	Page
Title and abstract				
Title	1	D;V	Identify the study as developing and/or validating a multivariable prediction model, the target population, and the outcome to be predicted.	1
Abstract	2	D;V	Provide a summary of objectives, study design, setting, participants, sample size, predictors, outcome, statistical analysis, results, and conclusions.	4-5
Introduction				
Background and objectives	3a	D;V	Explain the medical context (including whether diagnostic or prognostic) and rationale for developing or validating the multivariable prediction model, including references to existing models.	7-8
	3b	D;V	Specify the objectives, including whether the study describes the development or validation of the model or both.	7-9
Methods				
Source of data	4a	D;V	Describe the study design or source of data (e.g., randomized trial, cohort, or registry data), separately for the development and validation data sets, if applicable.	10
	4b	D;V	Specify the key study dates, including start of accrual; end of accrual; and, if applicable, end of follow-up.	10-11
Participants	5a	D;V	Specify key elements of the study setting (e.g., primary care, secondary care, general population) including number and location of centres.	11-12
	5b	D;V	Describe eligibility criteria for participants.	10-11
	5c	D;V	Give details of treatments received, if relevant.	10
Outcome	6a	D;V	Clearly define the outcome that is predicted by the prediction model, including how and when assessed.	12
	6b	D;V	Report any actions to blind assessment of the outcome to be predicted.	12
Predictors	7a	D;V	Clearly define all predictors used in developing or validating the multivariable prediction model, including how and when they were measured.	12-14
	7b	D;V	Report any actions to blind assessment of predictors for the outcome and other predictors.	14
Sample size	8	D;V	Explain how the study size was arrived at.	18
Missing data	9	D;V	Describe how missing data were handled (e.g., complete-case analysis, single imputation, multiple imputation) with details of any imputation method.	15
Statistical analysis methods	10a	D	Describe how predictors were handled in the analyses.	15-16
	10b	D	Specify type of model, all model-building procedures (including any predictor selection), and method for internal validation.	15-18
	10c	V	For validation, describe how the predictions were calculated.	18
	10d	D;V	Specify all measures used to assess model performance and, if relevant, to compare multiple models.	17
	10e	V	Describe any model updating (e.g., recalibration) arising from the validation, if done.	-
Risk groups	11	D;V	Provide details on how risk groups were created, if done.	-
Development vs. validation	12	V	For validation, identify any differences from the development data in setting, eligibility criteria, outcome, and predictors.	17-18
Results				
Participants	13a	D;V	Describe the flow of participants through the study, including the number of participants with and without the outcome and, if applicable, a summary of the follow-up time. A diagram may be helpful.	Figure 1
	13b	D;V	Describe the characteristics of the participants (basic demographics, clinical features, available predictors), including the number of participants with missing data for predictors and outcome.	-
	13c	V	For validation, show a comparison with the development data of the distribution of important variables (demographics, predictors and outcome).	-
Model development	14a	D	Specify the number of participants and outcome events in each analysis.	-
	14b	D	If done, report the unadjusted association between each candidate predictor and outcome.	-
Model	15a	D	Present the full prediction model to allow predictions for individuals (i.e., all regression	-

For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

TRIPOD Checklist: Prediction Model Development and Validation

specification			coefficients, and model intercept or baseline survival at a given time point).	-
	15b	D	Explain how to the use the prediction model.	-
Model performance	16	D;V	Report performance measures (with CIs) for the prediction model.	-
Model-updating	17	V	If done, report the results from any model updating (i.e., model specification, model performance).	-
Discussion				
Limitations	18	D;V	Discuss any limitations of the study (such as nonrepresentative sample, few events per predictor, missing data).	21
Interpretation	19a	V	For validation, discuss the results with reference to performance in the development data, and any other validation data.	-
	19b	D;V	Give an overall interpretation of the results, considering objectives, limitations, results from similar studies, and other relevant evidence.	-
Implications	20	D;V	Discuss the potential clinical use of the model and implications for future research.	21-22
Other information				
Supplementary information	21	D;V	Provide information about the availability of supplementary resources, such as study protocol, Web calculator, and data sets.	-
Funding	22	D;V	Give the source of funding and the role of the funders for the present study.	2

*Items relevant only to the development of a prediction model are denoted by D, items relating solely to a validation of a prediction model are denoted by V, and items relating to both are denoted D;V. We recommend using the TRIPOD Checklist in conjunction with the TRIPOD Explanation and Elaboration document.

For peer review only

BMJ Open

Development of screening tools to predict the risk of recurrence and related complications following anal fistula surgery: protocol for a prospective cohort study

Journal:	*BMJ Open*
Manuscript ID	bmjopen-2019-035134.R2
Article Type:	Protocol
Date Submitted by the Author:	04-Feb-2020
Complete List of Authors:	Mei, Zubing; Shuguang Hospital, Department of Anorectal Surgery Li, Yue; Shuguang Hospital, Department of Anorectal Surgery Zhang, Zhijun; Shuguang Hospital, Department of Anorectal Surgery Zhou, Haikun; Shuguang Hospital, Department of Anorectal Surgery Liu, Suzhi; Shuguang Hospital, Department of Anorectal Surgery Han, Ye; Shuguang Hospital, Department of Anorectal Surgery Du, Peixin; Shuguang Hospital, Department of Anorectal Surgery Qin, Xiufang; Shuguang Hospital, Department of Nursing Shao, Zhuo; Changhai Hospital, Department of General Surgery Ge, Maojun; Shuguang Hospital, Department of General Surgery Wang, Qingming; Shuguang Hospital, Department of Anorectal Surgery Yang, Wei; Shuguang Hospital, Department of Anorectal Surgery
Primary Subject Heading:	Surgery
Secondary Subject Heading:	Medical management
Keywords:	SURGERY, Risk management < HEALTH SERVICES ADMINISTRATION & MANAGEMENT, Colorectal surgery < SURGERY, Protocols & guidelines < HEALTH SERVICES ADMINISTRATION & MANAGEMENT

SCHOLARONE™
Manuscripts

BMJ Open: first published as 10.1136/bmjopen-2019-035134 on 4 March 2020. Downloaded from http://bmjopen.bmj.com/ on April 20, 2024 by guest. Protected by copyright.

I, the Submitting Author has the right to grant and does grant on behalf of all authors of the Work (as defined in the below author licence), an exclusive licence and/or a non-exclusive licence for contributions from authors who are: i) UK Crown employees; ii) where BMJ has agreed a CC-BY licence shall apply, and/or iii) in accordance with the terms applicable for US Federal Government officers or employees acting as part of their official duties; on a worldwide, perpetual, irrevocable, royalty-free basis to BMJ Publishing Group Ltd ("BMJ") its licensees and where the relevant Journal is co-owned by BMJ to the co-owners of the Journal, to publish the Work in this journal and any other BMJ products and to exploit all rights, as set out in our licence.

The Submitting Author accepts and understands that any supply made under these terms is made by BMJ to the Submitting Author unless you are acting as an employee on behalf of your employer or a postgraduate student of an affiliated institution which is paying any applicable article publishing charge ("APC") for Open Access articles. Where the Submitting Author wishes to make the Work available on an Open Access basis (and intends to pay the relevant APC), the terms of reuse of such Open Access shall be governed by a Creative Commons licence – details of these licences and which Creative Commons licence will apply to this Work are set out in our licence referred to above.

Other than as permitted in any relevant BMJ Author's Self Archiving Policies, I confirm this Work has not been accepted for publication elsewhere, is not being considered for publication elsewhere and does not duplicate material already published. I confirm all authors consent to publication of this Work and authorise the granting of this licence.

For peer review only

Development of screening tools to predict the risk of recurrence and related complications following anal fistula surgery: protocol for a prospective cohort study

Zubing Mei[1, 2]*; Yue Li[1]*; Zhijun Zhang[1]; Haikun Zhou[1]; Suzhi Liu[1]; Ye Han[1]; Peixin Du[1]; Xiufang Qin[3], Zhuo Shao[4]; Maojun Ge[5]; Qingming Wang [1]; Wei Yang[1, 2]

*Contributed equally to this manuscript

1. Department of Anorectal Surgery, Shuguang Hospital, Shanghai University of Traditional Chinese Medicine, Shanghai, China;
2. Anorectal Disease Institute of Shuguang Hospital, Shanghai, China;
3. Department of Nursing, Shuguang Hospital, Shanghai University of Traditional Chinese Medicine, Shanghai, China;
4. Department of General Surgery, Changhai Hospital, The Second Military Medical University, Shanghai, China;
5. Department of General Surgery, Shuguang Hospital, Shanghai University of Traditional Chinese Medicine, Shanghai, China.

Correspondence to:

Dr. Zubing Mei, MD, PhD

Department of Anorectal Surgery, Shuguang Hospital, Shanghai University of Traditional Chinese Medicine; Anorectal Disease Institute of Shuguang Hospital

528 Zhangheng Road, Shanghai 201203, China

Tel: 086-2120256187

E-mail: herrmayor@126.com

Professor. Wei Yang, MD

Department of Anorectal Surgery, Shuguang Hospital, Shanghai University of Traditional Chinese Medicine; Anorectal Disease Institute of Shuguang Hospital

528 Zhangheng Road, Shanghai 201203, China

For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

Tel: 086-2120256187

E-mail: yangweiyishi@163.com

Running head: Risk prediction models for AFR and related complications

Abstract Word Count: 300

Word Count (excluding abstract): 3,709

References: 53

Figures: 1

Key words: anal fistula, recurrence, prediction model, complication, surgery, cohort study

Funding statement: This work was supported by the National Natural Science Foundation of China (grant no. 81774112, 81673768), a grant from Siming Scholars from Shuguang Hospital (grant no. SGXZ-201913), Science and Technology Commission of Shanghai Municipality (grant no. 18401904600, 19401933900 and 16401971600), TCM dominant disease breeding project (grant no. zybz-2017018), the Plateau Peak Disciplines Innovation Team Project of Shanghai Municipal Education Commission (grant no. A1-N1920501010220), Shanghai Clinical Key Specialty (Chinese Medicine Anorectal Disease) Construction Foundation of the 13th Five-Year Plan (grant no. shslczdzk04302), Shanghai Famous Old Traditional Chinese Medicine Experts Inheritance Studio Construction Program (grant no.SHGZS-2017028),

For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

Famous Old Traditional Chinese Medicine Experts Inheritance Studio Construction Program of Shanghai University of TCM (grant no.SZYMZYGZS4016-2017028).

Abbreviations: AF, anal fistula; AFR, anal fistula recurrence; BMI, body mass index; FCS, fully conditional specification; MVNI, multivariate normal imputation; TRIPOD, Transparent Reporting of a multivariable prediction model for Individual Prognosis Or Diagnosis.

Author contributions:

Dr. Zubing Mei had full access to all of the data in the study and takes responsibility for the integrity of the data and the accuracy of the data analysis. Drs Mei and Li are co-first authors of this article.

Study concept and design: Zubing Mei.

Acquisition, analysis, or interpretation of data: Yue Li, Zubing Mei, Zhijun Zhang, Ye Han, Suzhi Liu, Haikun Zhou, Peixin Du, Xiufang Qin, Qingming Wang, Wei Yang.

Drafting of the manuscript: Zubing Mei.

Critical revision of the manuscript for important intellectual content: All authors.

Statistical analysis: Zubing Mei, Zhuo Shao, Maojun Ge.

Administrative, technical, or material support: All authors.

Study supervision: Zubing Mei.

Competing interests statement:

For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

The authors declare that they have no conflict of interest.

Role of the Funder/Sponsor:

The funder of the study had no role in the study design, data collection, data analysis, data interpretation or writing of the manuscript. The corresponding author had full access to all the data in the study and has final responsibility for the decision to submit for publication.

For peer review only

Abstract

Introduction

Postoperative recurrence and related complications are common and related to poor outcomes in patients with anal fistula (AF). Due to being associated with short-term and long-term cure rates, perioperative complications have received widespread attention following AF surgery. This study aims to identify a set of predictive factors to develop risk prediction models for recurrence and related complications following AF surgery. We plan to develop and validate risk prediction models, using information collected through a WeChat patient-reported questionnaire system combined with clinical, laboratory and imaging findings from the perioperative period until 3-6 months following AF surgery.

Methods and analysis

This is a prospective hospital-based cohort study using a linked database of collected health data as well as the follow-up outcomes for all adult patients who suffered from AF at a tertiary referral hospital in Shanghai, China. We will perform logistic regression models to predict anal fistula recurrence (AFR) as well as related complications (e.g., wound haemorrhage, faecal impaction, urinary retention, delayed wound healing and unplanned hospitalization) during and after AF surgery, and machine learning approaches will also be applied to develop risk prediction models. This prospective study aims to develop the first risk prediction models for AFR and related complications using multidimensional variables. These tools can be used to

For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

warn, motivate and empower patients to avoid some modifiable risk factors to prevent postoperative complications early. This study will also provide alternative tools for the early screening of high-risk patients with AFR and related complications, helping surgeons better understand the aetiology and outcomes of AF in an earlier stage.

Ethics and dissemination

The study was approved by the Institutional Review Board of Shuguang Hospital affiliated with Shanghai University of Traditional Chinese Medicine (approval number: 2019-699-54-01). The results of this study will be submitted to international scientific peer-reviewed journals or conferences in surgery, anorectal surgery or anorectal diseases.

Trial registration number

Chinese Clinical Trial Registry (ChiCTR1900025069); Pre-results.

Key words

Anal fistula; treatment outcome; recurrence; surgery; cohort study; prediction model

For peer review only

For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

Strengths and limitations of this study

- This is the first large prospective cohort study of patients with anal fistula at a tertiary referral hospital in China.
- A higher events per candidate predictive variable (≥20) will be applied which can generally eliminate bias in regression coefficients for prediction models and guarantee a sufficient sample size for model development.
- Candidate predictors will be identified from published and updated systematic reviews, expert opinions from Delphi surveys, and univariable or multivariable logistic regression analysis.
- The bootstrapping procedure will be applied for the internal and external validation of the prediction models.
- A higher probability of missing data due to non-response bias may occur as many of the variables are collected through a WeChat questionnaire system.

For peer review only

For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

Introduction

Anal fistula (AF) is a common perianal condition defined by a pathological epithelial tract that connects the anal canal or rectum and the surface of the perianal region, which is also regarded as a chronic stage of perianal abscess.[1] Postoperative recurrence, defined as persistence or recurrence of AF symptoms, or the development of recurrent perianal sepsis or chronic AF within six months of surgery,[2 3] is not only one of the consequences that can be related to a poorly performed surgical procedure but may also be due to the insidiousness of the disease. Our recently published meta-analysis based on 20 studies reported a recurrence rate of approximately 19% (95% CI 0.15-0.23) in patients undergoing AF surgery.[4] Because of the high degree of difficulty of surgery for patients with high complex AF, the postoperative recurrence rate of these patients can be as high as 50%, and the failure rate of reoperation remains at 10%.[5-8] It is considered one of the most difficult and complicated anorectal diseases.

A large number of studies have shown that the recurrence of AF is related to multiple factors, such as unclear diagnosis or failure to focus on the correct internal orifice, blind stump of fistula, incorrect method of seton, omission of branch of the fistula and poor drainage.[9 10] Studies also reported that anal fistula recurrence (AFR) was associated with individual patient characteristics, such as history of enteritis, previous anal surgery, obesity and smoking.[4 11-13] Li et.al retrospectively analysed 1783 patients with AF receiving surgical treatment and found that the location of AF,

For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

previous perianal surgery, seton history and enteritis were independent risk factors for AFR.[5] Recently, according to the evidence grading criteria based on Egger's P value, total sample size and between-study heterogeneity, we published a meta-analysis involving 20 studies with 6168 patients and concluded that high transsphincteric fistula, unidentified internal opening, and horseshoe extensions were independent risk factors for AFR with high-quality evidence, while prior anal surgery, seton placement surgery, and multiple fistula tracts were demonstrated to be risk factors for AFR with moderate-quality evidence.[4]

Factors influencing other perioperative complications related AF surgery including wound haemorrhage, faecal impaction, urinary retention, delayed wound healing and unplanned hospitalization are also rarely reported. Therefore, there is an urgent need to develop risk prediction tools for the complete profile of risk factors for AFR and related complications.

Which AF patients will be cured after surgery and which ones will not, are rarely investigated. The development of a prediction model for AFR following surgery to identify those patients with a higher risk of developing complications during follow up, would be of significant importance. First, surgeons can provide patients preoperatively with an estimated surgical cure rate according to the prediction models. Moreover, the current knowledge in the literature reporting potential predictive factors could help patients become familiar their individual risk factors and

For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

avoid modifiable ones to improve the cure rate, which has been well described and applied in the prevention of other diseases. [14-17]

However, to date, there are no effective screening tools to evaluate and predict the risk of recurrence or other adverse outcomes of AF. Therefore, the aim of the current study was to develop and validate multivariable prediction models that predict postoperative AFR and related complications.

Aims and objectives

The aim of this study was to develop risk prediction models for postoperative recurrence as well as other surgery-related complications in a prospective hospital-based AF cohort. A risk prediction model for perioperative complications will also be developed. A flowchart of prediction model development and assessment is provided in Figure 1.

The detailed tasks of this study are as follows:

1. Calculate the 3- to 6-month incidence of recurrence, and related complications in patients following AF surgery.

2. Establish the risk factors that significantly predict postoperative AFR and related complications based on the AF cohort in a tertiary referral centre.

3. Develop and validate the risk prediction models for postoperative AFR and related complications.

For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

4. Considering the different scenarios for different surgical interventions, conduct stratified analyses based on surgery type. If possible, a risk prediction model will also be developed in relevant subpopulations, such as those only receiving fistulectomy or fistulotomy, which can account for more than 60% of our AF cohort populations.

We also examine the following two hypotheses:

1. Patient-related demographic characteristics, fistula and surgery-related factors are predictive of postoperative AFR and related complications as dependent variables.

2. The risk prediction models for postoperative AFR and related complications developed in our study have more than 70% discriminating power.

Patients and Methods

Study design and participants

This study is a single-centre, prospective observational study on a hospital-based cohort enrolled at a tertiary referral centre in Shanghai, China.

Eligibility criteria

The enrolment of the cohort subjects was initiated in June, 2019. All subjects who will undergo surgical intervention for AF will be included. All operations will be performed by a group of colon and rectal surgeons at Shuguang Hospital, a regional tertiary referral centre. The exclusion criteria were age < 18 years, noncryptoglandular

For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

fistula (e.g., anal fistula due to inflammatory bowel disease, human immunodeficiency virus, malignant cancer, or obstetrical trauma), and rectovaginal or rectourethral fistula. The electronic medical records of the included subjects were complete.

Trained clinical investigators are collecting data in several categories, including baseline demographics, laboratory examinations, clinical data, imaging findings and follow-up information 3 to 6 months postoperatively. Planned clinical reviews or electronic surveys are conducted during hospitalization and every 0.5 to 3 months after discharge for 6 months.

Data collection

The research team comprised a principal investigator and 5 to 8 anorectal surgeons who were trained and supervised by the Ethics Committee of Shuguang Hospital. Written informed consent was obtained from all patients. The investigators did not intervene in any aspects of patient surveys at any stage of data collection and follow-up. Data were collected using a convenient follow-up system supported by Empower EDC (Solutions, Boston, Massachusetts, USA). This electronic system introduces a machine learning algorithm, through which we can use the data already entered in the Empower system to train the algorithm model and let the system itself develop quality control algorithms, validate the entered data and identify missing or suspicious data. Finally, the data manager will check the missing or suspicious data,

For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

confirm their completeness and ask the data manager to provide additional data when necessary. Furthermore, an automatic reminder follow-up function also plays a pivotal role during the whole follow-up period.

Patient and public involvement

Patients and the public will not be involved in the development, design, conduct or reporting of the study. The general results will be disseminated to participants through public education during follow-up.

Clinical outcomes

The primary study endpoint is postoperative recurrence following AF surgery defined as the persistence or recurrence of AF symptoms, or the development of recurrent perianal sepsis or chronic AF within 3 to 6 months of surgery.[2 3 18] The second end point is a composite outcome of postoperative comorbidities or any equivalent events including AFR, wound haemorrhage, faecal impaction, urinary retention, delayed wound healing or unplanned hospitalization associated with AF surgery. Outcomes were ascertained by the treating clinician combined with outpatient medical records or patient self-reports.

Selection of predictor variables

Candidate variables for the prediction model of the composite outcome of postoperative comorbidities will be screened according to the following pre-set

For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

criteria: (1) prior clinical knowledge; (2) results from a systematic review updated in November 2019 based on our published one [4] with sufficient evidence to include them as predictive variables in the risk model for AFR as is demonstrated below; or (3) agreed upon by a group of anorectal surgeons or experts for their clinical relevance using a two-round Delphi survey. We initially identified the following covariates as relevant candidate variables based on systematic reviews as well as clinical knowledge and/or relevance. The determination of all other candidate variables is based on the results of post-hoc analysis using univariable or multivariable survival analyses with a threshold of $p < 0.05$.

Factors identified from the systematic reviews and Delphi survey (manuscript under review), will be measured at baseline. These include factors involving the identified significant risk factors that are reported in our meta-analysis and are presented as follows:

- ► Prior anal surgery.
- ► Seton placement surgery.
- ► High transsphincteric fistula.
- ► Internal opening unidentified.
- ► Horseshoe extensions.
- ► Multiple fistula tracts.

For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

Some of the demographic factors and surgery details will also be collected due to limited power in the literature reviews and some non-significant potential factors (e.g., smoking or alcohol use) may be risk factors and are also included as follows:

► Gender.

► Age.

► Smoking.

► Alcohol use.

► Diabetes mellitus.

► Obesity.

► Preoperative seton drainage.

►High internal opening.

►Postoperative drainage.

►Supralevator extensions.

Data on other factors, such as laboratory examinations and MR imaging parameters (height of the internal openings, height and number of fistula, etc.) will also be collected. Moreover, other factors, such as chronic steroid therapy, diverting stoma, the surgeon's level of training, postoperative bowel confinement and antibiotic prophylaxis reported in previous literature, were selected for regression analysis as well. In addition, relevant factors from the expert-opinion survey were also assessed including the number of prior anal fistula surgeries, the types of surgery performed (such as staged fistulotomies, endorectal advancement flap and ligation of the

For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

intersphinteric fistula tract with and without seton drains), some nutrition parameters and immunomodulation medication use.

Categorization of potential predictors

We can code categorical predictors as "factor" variables, coding them as dummy variables. For example, smoking is coded originally as "1" for never smoker, "2" for past smoker, and "3" for current smoker and never smoker was selected as the reference category. A similar approach can be used with alcohol use.[19] Continuous variables formally should be measured with an interval or ratio scale, and should be able to take any value in a range. We treat ordered variables as linear which is generally reasonable for prediction. In other cases, continuous predictors can be grouped with meaningful categorization; for example, body mass index (BMI) can be classified based on internationally recognized categories (i.e., underweight, normal weight, overweight, and obesity).[20] Based on previous experiences, we will derive some predictors based on the responses of the surveys. However, in case some subjectivity in the classifications of these predictors may occur, sensitivity analyses will be performed to examine the robustness of our definitions during model development and validation.

Study quality control for the prediction models

Based on the summary of methodological quality and the developmental stage of prediction models by van Oort et al., we are developing predesigned criteria for the

For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

quality control of our prediction models, which can allow us to conduct the study and report the results more rigorously.[21 22] The methodological checklist of the study is presented in the supplementary material.

Missing data

Candidate predictors with more than 60% missingness will be excluded. For those with less than 60% missingness, multiple imputation are to be performed by imputing 20 complete data sets using multivariate normal regression, [23-26] which can reasonably approximate the true distributional relationship between the missing values and the available ones.[27] Among various multiple imputation approaches, fully conditional specification (FCS) and multivariate normal imputation (MVNI) are preferred, because they have been proven to be generally less biased than complete-case analysis. They can both generate similar results in the presence of either binary or ordinal variables that are not generally normally distributed. [27]

Statistical analysis for model derivation

Logistic regression will be applied to develop our prediction models for the binary outcomes. All data processing and statistical analysis will be performed using EmpowerStats software (www. empowerstats.com; X&Y Solutions, Inc., Boston, MA, USA) and the statistical software package R (R Foundation, Vienna, Austria).

We will first study the association between each potential variable and the outcome based on univariable analysis. Variables are considered further for multivariable

For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

regression modelling when they are associated with a p-value less than 0.20. Normality or linearity will be evaluated for the continuous predictors. Fractional polynomials are advocated for associations between the continuous predictors and the outcome for nonlinear relationships. [28 29] We will perform backward stepwise selection with $p < 0.001$ as the inclusion threshold and $p > 0.05$ as the exclusion threshold for each imputed data set.

Predictors that appear in the imputation models with an inclusion fraction of ≥50% are qualified for the final multivariable model. Although there is no consensus regarding the optimal method for selecting predictors for inclusion, backwards elimination is generally considered as the preferred procedure as reported by Mantel et al.[30] A forward stepwise procedure will also be performed to repeat the analysis to test the robustness of the models. Overall regression coefficient estimates of the models will be generated with the combination of the imputed datasets based on Rubin's Rules, while taking into account uncertainty in the imputed values.[23 24 26] Collinearity will also be assessed which refers to the fact that predictors can have strong correlation with each other, defined as correlation coefficient >0.8, or variance inflation factor >10.[31] Then we will examine the interactions among the regression models.

Prediction model performance assessment

For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

Prediction models will be developed with a random sample of 60% of the AF cohort as the derivation cohort, and then validated with the remaining sample of 40% of the cohort as the validation cohort. The predictive performance in the derivation and validation cohort will be evaluated and reported by examining measures of predictive accuracy, discrimination and calibration. Nagelkerke's R^2 and the Brier score will be used for the measurement of predictive accuracy.[32 33] The discriminative ability of the prediction models is evaluated using several statistics, which are according to the discriminative and calibration ability in both the derivation and validation AF cohorts. Model discrimination is the ability of the models to differentiate between high-risk patients and low-risk patients (having high or low risk of AFR or surgery-related complications). This will be assessed via Harrell's concordance statistic (C-index).[34] The calculation of the C-index will be performed in each of the 20 imputed data sets, and then averaged based on Rubin's rule.[35] The model is interpreted as having no discriminatory ability when a value of C-index is 0.5, and has perfect discrimination when a value of the C-index is 1.0.[34] Calibration implies the agreement between the predicted outcomes and the observed outcomes, which is evaluated with the Hosmer-Lemeshow goodness-of-fit test in all imputed datasets presented with calibration plots.[36] Calibration-in-the-large, which is defined as the agreement between mean observed outcomes and mean predictions, will also be assessed for calibration.[37]

Internal and external validation of the prediction model

For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

To make the prediction models reproducible, we must perform internal validation. The bootstrapping technique, as one of the most attractive resampling techniques, is a mostly applied validation method that seems to be most efficient for obtaining stable optimism-corrected estimates.[34 38] It has been reported that bootstrap validation is a feasible technique for most prediction models with at least a 500 bootstrap resampling procedures using Harrell's validation function, which can adjust the developed models for overfitting.[39] We will also apply temporal validation as external validation using a more recent AF patient cohort. [40]

Sample size

Since there are no widely accepted methods for the estimation of the sample size requirements to develop the risk prediction models, the size of this AF cohort will be calculated to have 20 events per candidate predictive variable (EPV, defined as the ratio of the number of individuals with the outcome event to the number of candidate predictors), which can generally eliminate bias in regression coefficients for prediction models with low-prevalence binary predictor development (the estimated recurrence rate <20%) and adequately power the logistic regression models.[41,42] According to the findings by Ogundimu et al. [43], a higher EPV (≥20) can generally eliminate bias in regression coefficients for prediction models with low-prevalence binary predictor development. Then we estimate 400 events allowing for 20 predictor variables (EPV=20). Considering a 5-20% recurrence rate, we assume that at least 4000-8000 patients should be recruited for model development. In addition, surgery

For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

type and fistula type-stratified analyses will also be performed to examine the different effects of these factors on disease recurrence or other related complications in each subgroup. Risk prediction models, if possible, can also be developed in those subpopulations. A cohort size with more than 4000-8000 patients will provide sufficient power to perform those analyses and develop prediction models in those subgroups.

Follow-up and methodological quality control

The application of WeChat questionnaires to collect data will inevitably increase the probability of missing data. However, we have made some predesigned countermeasures. For example, we have set up follow-up reminders via the WeChat questionnaire system. Moreover, every week two trained clinical fellows cross-check the data, and will contact the respondents by phone or WeChat about the missing contents, which can minimize the missing data and loss to follow-up rate.

A limited number of studies have identified specific criteria for quality control in a prediction model, but we have strictly adhered to the guidelines for the reporting of developing studies, validating multivariable clinical prediction models as is reported in the TRIPOD (Transparent Reporting of a multivariable prediction model for Individual Prognosis Or Diagnosis) Statement to ensure methodological rigour.[44] All issues have been addressed in this study in the supplementary material.

For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

Ethics and dissemination

The study was approved by the Institutional Review Board of Shuguang Hospital affiliated with Shanghai University of Traditional Chinese Medicine (approval number: 2019-699-54-01). The results of this cohort will be submitted to international scientific peer-reviewed journals or conferences in surgery, anorectal surgery or anorectal diseases.

Discussion

In this study, we plan to develop internally validated models for the prediction of the recurrrence as well as postoperative complications among AF patients. The models will be developed based on a large AF cohort in a hospital-representative linked database with validated clinical information. The collected variables include WeChat questionnaires, clinical, laboratory and imaging findings, and follow-up information, all of which are routinely being gathered at the time of enrolment.

According to the existing knowledge and systematic reviews, it is highly plausible that a number of patient, fistula and surgery related characteristics (e.g., patient-related variables such as gender, age, diabetes mellitus or obesity,[45-47] lifestyle factors such as smoking and alcohol abuse;[48 49] fistula-related factors such as the number of fistula tracts, horseshoe extensions, classification and location of fistula; and surgery-related variables such as prior anal surgery and postoperative drainage[48 50 51]) are easily ascertainable before surgery may predict AFR. Similar risk prediction

For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

models exist in other diseases, such as the Framingham risk score model to predict cardiovascular disease risk[52] and the Korean Crohn's disease prediction (KCDP) model to predict the clinical course of Crohn's disease.[53] Until now, risk factor investigations of predictors of perioperative surgery-related complications have been limited to assessments of single predictors with small sample sizes. Risk prediction models can help inform surgeons regarding high-risk AF patients based on the overall risk factors. The primary purpose of this study was to develop two risk prediction models to assist surgeons in identifying AF patients scheduled for surgical treatment who are at higher risk of developing recurrence and surgery-related complications. The predictive models will help both clinicians and patients identify the risk of complications after AF surgery in advance, and perform the interventions necessary to reduce the risk of surgery-related complications and the personal and social financial burden brought about by those complications. Accurate risk prediction models are especially instructive for the development of the optimal surgical plan to achieve optimal surgical outcome.

Our study has several strengths. To the best of our knowledge, this is the first study with the primary aim of developing and internally and externally validating multivariable prediction models for AFR and related complications following AF surgery. Multidimensional clinically useful candidate predictors will be fully examined from a variety of sources including our published and updated systematic reviews, expert opinions from Delphi surveys and univariable or multivariable logistic

For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

regression analysis. Second, we will apply the internal and external validation of the prediction models using the bootstrapping procedure. Third, multiple imputation will also be used to treat the missing data. Finally, our study is a prospective cohort study with an adequate follow-up period which can minimize certain forms of bias.

Our study also has limitations. As many of the variables are collected through a WeChat questionnaire system during hospitalization and follow-up, a higher probability of missing data due to non-response bias may occur. To address this issue, we regularly send reminders to those who do not respond after discharge. Second, although we will investigate a series of potential predictors, some more potential predictors will not be involved or not collected in the current study, such as data related to postoperative nursing strategy and outpatient follow-up frequency. Moreover, bias may also result from the single-centre recruitment of our study and will be improved through multicentre recruitment in the future. Finally, one key potential issue that needs to be considered a priori is the variable recurrence rate depending on the risk factors included and identified in the testing and validating cohorts since this may affect the C-index. It is possible that a lack of key variable inclusions in the models may result in decreased discriminatory ability. This is a function of the database and points of interest included that may need to be maximized before proceeding with development of the testing model.

For peer review only

For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

The newly developed risk algorithms may have significant applications in clinical practice by helping recommend an optimal surgical approach for a specific AF patient, as well as intensive perioperative care, education, and the timely assessment and discussion of the need for interventions among those most at the highest risk of developing recurrence or surgery-related complications. The models will specifically identify AF patients who are likely to develop recurrence or related complications following AF surgery to offer a quantitative evaluation of the risk. Moreover, the models will also provide reference information for preventing recurrence, reducing the rate of recurrence after operation and intervening in high-risk factors in the early stage.

In summary, this study protocol summarizes the design of development and validation studies for a risk screening tool in patients receiving AF surgery. The results from this study will be interpreted for the purpose of clinical decision making. The models to be developed in this study could be used to make new recommendations for perioperative AF patients.

For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

References

1. Tantiphlachiva K. 7 - Anorectal Abscess and Fistula. In: Coss-Adame E, Remes-Troche JM, eds. Anorectal Disorders: Academic Press 2019:81-100.

2. Abbas MA, Jackson CH, Haigh PI. Predictors of outcome for anal fistula surgery. Archives of surgery (Chicago, Ill : 1920) 2011;146(9):1011-6. doi: 10.1001/archsurg.2011.197

3. Hamadani A, Haigh P, Liu I, et al. Who is at risk for developing chronic anal fistula or recurrent anal sepsis after initial perianal abscess? Dis Colon Rectum 2009;52(2):217-21.

4. Mei Z, Wang Q, Zhang Y, et al. Risk Factors for Recurrence after anal fistula surgery: A meta-analysis. International Journal of Surgery 2019;69:153-64. doi: https://doi.org/10.1016/j.ijsu.2019.08.003

5. Li J, Yang W, Huang Z, et al. [Clinical characteristics and risk factors for recurrence of anal fistula patients]. Zhonghua wei chang wai ke za zhi = Chinese journal of gastrointestinal surgery 2016;19(12):1370-74.

6. Gottgens KW, Smeets RR, Stassen LP, et al. Systematic review and meta-analysis of surgical interventions for high cryptoglandular perianal fistula. International journal of colorectal disease 2015;30(5):583-93. doi: 10.1007/s00384-014-2091-8

For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

7. Gottgens KW, Janssen PT, Heemskerk J, et al. Long-term outcome of low perianal fistulas treated by fistulotomy: a multicenter study. International journal of colorectal disease 2015;30(2):213-9. doi: 10.1007/s00384-014-2072-y

8. Bakhtawar N, Usman M. Factors Increasing the Risk of Recurrence in Fistula-in-ano. Cureus 2019;11(3):e4200. doi: 10.7759/cureus.4200

9. Sygut A, Mik M, Trzcinski R, et al. How the location of the internal opening of anal fistulas affect the treatment results of primary transsphincteric fistulas. Langenbeck's archives of surgery 2010;395(8):1055-9. doi: 10.1007/s00423-009-0562-0

10. Dziki A, Bartos M. Seton treatment of anal fistula: experience with a new modification. The European journal of surgery = Acta chirurgica 1998;164(7):543-8. doi: 10.1080/110241598750005930

11. Hermann J, Eder P, Banasiewicz T, et al. Current management of anal fistulas in Crohn's disease. Przeglad gastroenterologiczny 2015;10(2):83-8. doi: 10.5114/pg.2015.49684

12. Hamadani A, Haigh PI, Liu IL, et al. Who is at risk for developing chronic anal fistula or recurrent anal sepsis after initial perianal abscess? Diseases of the colon and rectum 2009;52(2):217-21. doi: 10.1007/DCR.0b013e31819a5c52

For peer review only

13. Ellis CN, Clark S. Effect of tobacco smoking on advancement flap repair of complex anal fistulas. Diseases of the colon and rectum 2007;50(4):459-63. doi: 10.1007/s10350-006-0829-2

14. Moons KG, Altman DG, Vergouwe Y, et al. Prognosis and prognostic research: application and impact of prognostic models in clinical practice. Bmj 2009;338:b606. doi: 10.1136/bmj.b606

15. Ming C, Viassolo V, Probst-Hensch N, et al. Machine learning techniques for personalized breast cancer risk prediction: comparison with the BCRAT and BOADICEA models. Breast cancer research : BCR 2019;21(1):75. doi: 10.1186/s13058-019-1158-4

16. Khan SS, Ning H, Shah SJ, et al. 10-Year Risk Equations for Incident Heart Failure in the General Population. Journal of the American College of Cardiology 2019;73(19):2388-97. doi: 10.1016/j.jacc.2019.02.057

17. Terry MB, Liao Y, Whittemore AS, et al. 10-year performance of four models of breast cancer risk: a validation study. The Lancet Oncology 2019;20(4):504-17. doi: 10.1016/S1470-2045(18)30902-1

18. Yano T, Asano M, Matsuda Y, et al. Prognostic factors for recurrence following the initial drainage of an anorectal abscess. International journal of colorectal disease 2010;25(12):1495-8. doi: 10.1007/s00384-010-1011-9 [published Online First: 2010/07/20]

For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

19. Steyerberg EW. Coding of Categorical and Continuous Predictors. In: Steyerberg EW, ed. Clinical Prediction Models: A Practical Approach to Development, Validation, and Updating. Cham: Springer International Publishing 2019:175-90.

20. Morgan TM, Elashoff RM. Effect of Categorizing a Continuous Covariate on the Comparison of Survival Time. Journal of the American Statistical Association 1986;81(396):917-21. doi: 10.1080/01621459.1986.10478350

21. van Oort L, van den Berg T, Koes BW, et al. Preliminary state of development of prediction models for primary care physical therapy: a systematic review. Journal of Clinical Epidemiology 2012;65(12):1257-66. doi: https://doi.org/10.1016/j.jclinepi.2012.05.007

22. Rushton AB, Verra ML, Emms A, et al. Development and validation of two clinical prediction models to inform clinical decision-making for lumbar spinal fusion surgery for degenerative disorders and rehabilitation following surgery: protocol for a prospective observational study. BMJ Open 2018;8(5):e021078. doi: 10.1136/bmjopen-2017-021078

23. Wood AM, White IR, Royston P. How should variable selection be performed with multiply imputed data? Statistics in Medicine 2008;27(17):3227-46. doi: 10.1002/sim.3177

For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

24. White IR, Royston P, Wood AM. Multiple imputation using chained equations: Issues and guidance for practice. Statistics in Medicine 2011;30(4):377-99. doi: 10.1002/sim.4067

25. Schafer JL. Multiple imputation: a primer. Statistical Methods in Medical Research 1999;8(1):3-15. doi: 10.1177/096228029900800102

26. Marshall A, Altman DG, Holder RL, et al. Combining estimates of interest in prognostic modelling studies after multiple imputation: current practice and guidelines. BMC Medical Research Methodology 2009;9(1):57. doi: 10.1186/1471-2288-9-57

27. Lee KJ, Carlin JB. Multiple Imputation for Missing Data: Fully Conditional Specification Versus Multivariate Normal Imputation. American Journal of Epidemiology 2010;171(5):624-32. doi: 10.1093/aje/kwp425

28. Sauerbrei W, Royston P, Binder H. Selection of important variables and determination of functional form for continuous predictors in multivariable model building. Statistics in Medicine 2007;26(30):5512-28. doi: 10.1002/sim.3148

29. Royston P, Sauerbrei W. Building multivariable regression models with continuous covariates in clinical epidemiology--with an emphasis on fractional polynomials. Methods of information in medicine 2005;44(4):561-71.

30. Mantel N. Why Stepdown Procedures in Variable Selection. Technometrics 1970;12(3):621-25. doi: 10.1080/00401706.1970.10488701

For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

31. Vittinghoff E, Glidden DV, Shiboski SC, et al. Regression methods in biostatistics: linear, logistic, survival, and repeated measures models: Springer Science & Business Media 2011.

32. Steyerberg EW, Vickers AJ, Cook NR, et al. Assessing the performance of prediction models: a framework for traditional and novel measures. Epidemiology 2010;21(1):128-38. doi: 10.1097/EDE.0b013e3181c30fb2

33. NAGELKERKE NJD. A note on a general definition of the coefficient of determination. Biometrika 1991;78(3):691-92. doi: 10.1093/biomet/78.3.691

34. Harrell Jr FE. Regression modeling strategies: with applications to linear models, logistic and ordinal regression, and survival analysis: Springer 2015.

35. Rubin DB. Multiple imputation for nonresponse in surveys: John Wiley & Sons 2004.

36. Hosmer Jr DW, Lemeshow S, Sturdivant RX. Applied logistic regression: John Wiley & Sons 2013.

37. Snell KI, Ensor J, Debray TP, et al. Meta-analysis of prediction model performance across multiple studies: Which scale helps ensure between-study normality for the C-statistic and calibration measures? Statistical methods in medical research 2018;27(11):3505-22.

For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

38. Chatfield C. Model uncertainty, data mining and statistical inference. Journal of the Royal Statistical Society: Series A (Statistics in Society) 1995;158(3):419-44.

39. Steyerberg EW, Bleeker SE, Moll HA, et al. Internal and external validation of predictive models: a simulation study of bias and precision in small samples. Journal of clinical epidemiology 2003;56(5):441-47.

40. Justice AC, Covinsky KE, Berlin JA. Assessing the generalizability of prognostic information. Annals of internal medicine 1999;130(6):515-24.

41. Concato J, Peduzzi P, Holford TR, et al. Importance of events per independent variable in proportional hazards analysis I. Background, goals, and general strategy. Journal of clinical epidemiology 1995;48(12):1495-501.

42. Steyerberg EW. Clinical Prediction Models: A Practical Approach to Development, Validation, and Updating. SPRINGER NATURE, 2019.

43. Ogundimu EO, Altman DG, Collins GS. Adequate sample size for developing prediction models is not simply related to events per variable. J Clin Epidemiol. 2016;76:175–182.

44. Moons KG, Altman DG, Reitsma JB, et al. Transparent Reporting of a multivariable prediction model for Individual Prognosis or Diagnosis (TRIPOD): explanation and elaboration. Annals of internal medicine 2015;162(1):W1-W73.

For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

45. Hamadani A, Haigh PI, Liu I-LA, et al. Who is at risk for developing chronic anal fistula or recurrent anal sepsis after initial perianal abscess? Diseases of the colon & rectum 2009;52(2):217-21.

46. Sonoda T, Hull T, Piedmonte MR, et al. Outcomes of primary repair of anorectal and rectovaginal fistulas using the endorectal advancement flap. Diseases of the colon & rectum 2002;45(12):1622-28.

47. Hyman N, O'Brien S, Osler T. Outcomes after fistulotomy: results of a prospective, multicenter regional study. Diseases of the Colon & Rectum 2009;52(12):2022-27.

48. Ellis CN, Clark S. Effect of tobacco smoking on advancement flap repair of complex anal fistulas. Diseases of the colon & rectum 2007;50(4):459-63.

49. Schwandner T, Roblick MH, Kierer W, et al. Surgical treatment of complex anal fistulas with the anal fistula plug: a prospective, multicenter study. Diseases of the Colon & Rectum 2009;52(9):1578-83.

50. Garcia-Aguilar J, Belmonte C, Wong WD, et al. Anal fistula surgery: Factors associated with recurrence and incontinence. Diseases of the colon & rectum 1996;39(7):723-29.

51. Athanasiadis S, Köhler A, Nafe M. Treatment of high anal fistulae by primary occlusion of the internal ostium, drainage of the intersphincteric space, and mucosal advancement flap. International journal of colorectal disease 1994;9(3):153-57.

For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

52. Fox ER, Samdarshi TE, Musani SK, et al. Development and validation of risk prediction models for cardiovascular events in black adults: the Jackson Heart Study Cohort. JAMA cardiology 2016;1(1):15-25.

53. Park Y, Cheon JH, Park YL, et al. Development of a novel predictive model for the clinical course of Crohn's disease: results from the connect study. Inflammatory bowel diseases 2017;23(7):1071-79.

For peer review only

For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

Figure legend

Figure 1. Flowchart of prediction model development and assessment.

For peer review only

For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

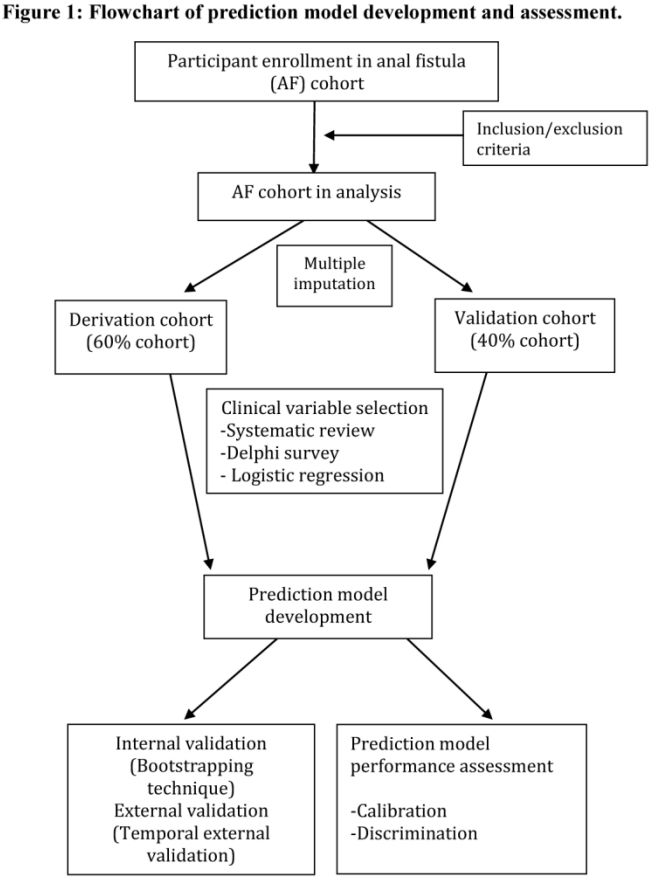

Figure 1. Flowchart of prediction model development and assessment.

BMJ Open: first published as 10.1136/bmjopen-2019-035134 on 4 March 2020. Downloaded from http://bmjopen.bmj.com/ on April 20, 2024 by guest. Protected by copyright.

TRIPOD Checklist: Prediction Model Development and Validation

This document contains a completed TRIPOD checklist for the manuscript: "**Development of screening tools to predict the risk of recurrence and related complications following anal fistula surgery: protocol for a prospective cohort study"**.

Zubing Mei, Yue Li, Zhijun Zhang, Haikun Zhou, Suzhi Liu, Ye Han, Peixin Du, Xiufang Qin, Zhuo Shao, Maojun Ge, Qingming Wang, Wei Yang.

As this TRIPOD checklist refers to a study protocol, not all items are relevant at this stage. We have tried to the furthest extent possible to make the protocol adhere to the TRIPOD checklist and all reporting of results will be in accordance to the protocol and the TRIPOD statement. Page numbers in the submitted manuscript are provided. For items that are only partly relevant at this time, page numbers are provided in parentheses and for items that are not relevant at this time a "-" has been written.

Section/Topic	Item		Checklist Item	Page
Title and abstract				
Title	1	D;V	Identify the study as developing and/or validating a multivariable prediction model, the target population, and the outcome to be predicted.	1
Abstract	2	D;V	Provide a summary of objectives, study design, setting, participants, sample size, predictors, outcome, statistical analysis, results, and conclusions.	5-6
Introduction				
Background and objectives	3a	D;V	Explain the medical context (including whether diagnostic or prognostic) and rationale for developing or validating the multivariable prediction model, including references to existing models.	8-9
	3b	D;V	Specify the objectives, including whether the study describes the development or validation of the model or both.	8-10
Methods				
Source of data	4a	D;V	Describe the study design or source of data (e.g., randomized trial, cohort, or registry data), separately for the development and validation data sets, if applicable.	11
	4b	D;V	Specify the key study dates, including start of accrual; end of accrual; and, if applicable, end of follow-up.	11-12
Participants	5a	D;V	Specify key elements of the study setting (e.g., primary care, secondary care, general population) including number and location of centres.	12-13
	5b	D;V	Describe eligibility criteria for participants.	11-12
	5c	D;V	Give details of treatments received, if relevant.	11
Outcome	6a	D;V	Clearly define the outcome that is predicted by the prediction model, including how and when assessed.	13
	6b	D;V	Report any actions to blind assessment of the outcome to be predicted.	13
Predictors	7a	D;V	Clearly define all predictors used in developing or validating the multivariable prediction model, including how and when they were measured.	13-15
	7b	D;V	Report any actions to blind assessment of predictors for the outcome and other predictors.	15
Sample size	8	D;V	Explain how the study size was arrived at.	20
Missing data	9	D;V	Describe how missing data were handled (e.g., complete-case analysis, single imputation, multiple imputation) with details of any imputation method.	17
Statistical analysis methods	10a	D	Describe how predictors were handled in the analyses.	17-18
	10b	D	Specify type of model, all model-building procedures (including any predictor selection), and method for internal validation.	17-20
	10c	V	For validation, describe how the predictions were calculated.	20
	10d	D;V	Specify all measures used to assess model performance and, if relevant, to compare multiple models.	19
	10e	V	Describe any model updating (e.g., recalibration) arising from the validation, if done.	-
Risk groups	11	D;V	Provide details on how risk groups were created, if done.	-
Development vs. validation	12	V	For validation, identify any differences from the development data in setting, eligibility criteria, outcome, and predictors.	19-20
Results				
Participants	13a	D;V	Describe the flow of participants through the study, including the number of participants with and without the outcome and, if applicable, a summary of the follow-up time. A diagram may be helpful.	Figure 1
	13b	D;V	Describe the characteristics of the participants (basic demographics, clinical features, available predictors), including the number of participants with missing data for predictors and outcome.	-
	13c	V	For validation, show a comparison with the development data of the distribution of important variables (demographics, predictors and outcome).	-
Model development	14a	D	Specify the number of participants and outcome events in each analysis.	-
	14b	D	If done, report the unadjusted association between each candidate predictor and outcome.	-
Model	15a	D	Present the full prediction model to allow predictions for individuals (i.e., all regression	-

For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

TRIPOD Checklist: Prediction Model Development and Validation

specification			coefficients, and model intercept or baseline survival at a given time point).	-
	15b	D	Explain how to the use the prediction model.	-
Model performance	16	D;V	Report performance measures (with CIs) for the prediction model.	-
Model-updating	17	V	If done, report the results from any model updating (i.e., model specification, model performance).	-
Discussion				
Limitations	18	D;V	Discuss any limitations of the study (such as nonrepresentative sample, few events per predictor, missing data).	24
Interpretation	19a	V	For validation, discuss the results with reference to performance in the development data, and any other validation data.	-
	19b	D;V	Give an overall interpretation of the results, considering objectives, limitations, results from similar studies, and other relevant evidence.	-
Implications	20	D;V	Discuss the potential clinical use of the model and implications for future research.	24-25
Other information				
Supplementary information	21	D;V	Provide information about the availability of supplementary resources, such as study protocol, Web calculator, and data sets.	-
Funding	22	D;V	Give the source of funding and the role of the funders for the present study.	2

*Items relevant only to the development of a prediction model are denoted by D, items relating solely to a validation of a prediction model are denoted by V, and items relating to both are denoted D;V. We recommend using the TRIPOD Checklist in conjunction with the TRIPOD Explanation and Elaboration document.

For peer review only

SPRINGER NATURE

Author Services

Editing Certificate

This document certifies that the manuscript

Development of screening tools to predict the risk of recurrence and related complications following anal fistula surgery: protocol for a prospective cohort study

prepared by the authors

Zubing Mei, Yue Li, Zhijun Zhang, Haikun Zhou, Suzhi Liu,Ye Han, Peixin Du, Xiufang Qin,Zhuo Shao,Maojun Ge, Qingming Wang ,Wei Yang

was edited for proper English language, grammar, punctuation, spelling, and overall style by one or more of the highly qualified native English speaking editors at SNAS.

This certificate was issued on **February 4, 2020** and may be verified on the SNAS website using the verification code **B8CF-ECA7-4F33-353A-DEE0**.

Neither the research content nor the authors' intentions were altered in any way during the editing process. Documents receiving this certification should be English-ready for publication; however, the author has the ability to accept or reject our suggestions and changes. To verify the final SNAS edited version, please visit our verification page at secure.authorservices.springernature.com/certificate/verify. If you have any questions or concerns about this edited document, please contact SNAS at support@as.springernature.com.

SNAS provides a range of editing, translation, and manuscript services for researchers and publishers around the world. For more information about our company, services, and partner discounts, please visit authorservices.springernature.com.

For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

1136/bmjopen-2019-035134 on 4 March 2020. Downloaded from http://bmjopen.bmj.com/ on April 20, 2024 by guest. Protected by copyright.

TRIPOD Checklist: Prediction Model Development and Validation

This document contains a completed TRIPOD checklist for the manuscript: "**Development of screening tools to predict the risk of recurrence and related complications following anal fistula surgery: a prospective cohort study"**.

Zubing Mei, Yue Li, Zhijun Zhang, Suzhi Liu, Haikun Zhou, Ye Han, Peixin Du, Zhuo Shao, Maojun Ge, Qingming Wang, Wei Yang.

As this TRIPOD checklist refers to a study protocol, not all items are relevant at this stage. We have tried to the furthest extent possible to make the protocol adhere to the TRIPOD checklist and all reporting of results will be in accordance to the protocol and the TRIPOD statement. Page numbers in the submitted manuscript are provided. For items that are only partly relevant at this time, page numbers are provided in parentheses and for items that are not relevant at this time a "-" has been written.

Section/Topic	Item		Checklist Item	Page
Title and abstract				
Title	1	D;V	Identify the study as developing and/or validating a multivariable prediction model, the target population, and the outcome to be predicted.	1
Abstract	2	D;V	Provide a summary of objectives, study design, setting, participants, sample size, predictors, outcome, statistical analysis, results, and conclusions.	4-5
Introduction				
Background and objectives	3a	D;V	Explain the medical context (including whether diagnostic or prognostic) and rationale for developing or validating the multivariable prediction model, including references to existing models.	7-8
	3b	D;V	Specify the objectives, including whether the study describes the development or validation of the model or both.	7-9
Methods				
Source of data	4a	D;V	Describe the study design or source of data (e.g., randomized trial, cohort, or registry data), separately for the development and validation data sets, if applicable.	10
	4b	D;V	Specify the key study dates, including start of accrual; end of accrual; and, if applicable, end of follow-up.	10-11
Participants	5a	D;V	Specify key elements of the study setting (e.g., primary care, secondary care, general population) including number and location of centres.	11-12
	5b	D;V	Describe eligibility criteria for participants.	10-11
	5c	D;V	Give details of treatments received, if relevant.	10
Outcome	6a	D;V	Clearly define the outcome that is predicted by the prediction model, including how and when assessed.	12
	6b	D;V	Report any actions to blind assessment of the outcome to be predicted.	12
Predictors	7a	D;V	Clearly define all predictors used in developing or validating the multivariable prediction model, including how and when they were measured.	12-14
	7b	D;V	Report any actions to blind assessment of predictors for the outcome and other predictors.	14
Sample size	8	D;V	Explain how the study size was arrived at.	18
Missing data	9	D;V	Describe how missing data were handled (e.g., complete-case analysis, single imputation, multiple imputation) with details of any imputation method.	15
Statistical analysis methods	10a	D	Describe how predictors were handled in the analyses.	15-16
	10b	D	Specify type of model, all model-building procedures (including any predictor selection), and method for internal validation.	15-18
	10c	V	For validation, describe how the predictions were calculated.	18
	10d	D;V	Specify all measures used to assess model performance and, if relevant, to compare multiple models.	17
	10e	V	Describe any model updating (e.g., recalibration) arising from the validation, if done.	-
Risk groups	11	D;V	Provide details on how risk groups were created, if done.	-
Development vs. validation	12	V	For validation, identify any differences from the development data in setting, eligibility criteria, outcome, and predictors.	17-18
Results				
Participants	13a	D;V	Describe the flow of participants through the study, including the number of participants with and without the outcome and, if applicable, a summary of the follow-up time. A diagram may be helpful.	Figure 1
	13b	D;V	Describe the characteristics of the participants (basic demographics, clinical features, available predictors), including the number of participants with missing data for predictors and outcome.	-
	13c	V	For validation, show a comparison with the development data of the distribution of important variables (demographics, predictors and outcome).	-
Model development	14a	D	Specify the number of participants and outcome events in each analysis.	-
	14b	D	If done, report the unadjusted association between each candidate predictor and outcome.	-
Model	15a	D	Present the full prediction model to allow predictions for individuals (i.e., all regression	-

For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

TRIPOD Checklist: Prediction Model Development and Validation

specification			coefficients, and model intercept or baseline survival at a given time point).	-
	15b	D	Explain how to the use the prediction model.	-
Model performance	16	D;V	Report performance measures (with CIs) for the prediction model.	-
Model-updating	17	V	If done, report the results from any model updating (i.e., model specification, model performance).	-
Discussion				
Limitations	18	D;V	Discuss any limitations of the study (such as nonrepresentative sample, few events per predictor, missing data).	21
Interpretation	19a	V	For validation, discuss the results with reference to performance in the development data, and any other validation data.	-
	19b	D;V	Give an overall interpretation of the results, considering objectives, limitations, results from similar studies, and other relevant evidence.	-
Implications	20	D;V	Discuss the potential clinical use of the model and implications for future research.	21-22
Other information				
Supplementary information	21	D;V	Provide information about the availability of supplementary resources, such as study protocol, Web calculator, and data sets.	-
Funding	22	D;V	Give the source of funding and the role of the funders for the present study.	2

*Items relevant only to the development of a prediction model are denoted by D, items relating solely to a validation of a prediction model are denoted by V, and items relating to both are denoted D;V. We recommend using the TRIPOD Checklist in conjunction with the TRIPOD Explanation and Elaboration document.

For peer review only