

# 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](mailto:info.bmjopen@bmj.com)

# BMJ Open

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

Journal:	<i>BMJ Open</i>
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



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.

1  
2  
3  
4 **Development of screening tools to predict the risk of recurrence and related**  
5 **complications following anal fistula surgery: a prospective cohort study**  
6  
7

8  
9 Zubing Mei<sup>1,2\*</sup>; Yue Li<sup>1\*</sup>; Zhijun Zhang<sup>1</sup>; Suzhi Liu<sup>1</sup>; Haikun Zhou<sup>1</sup>; Ye Han<sup>1</sup>;  
10 Peixin Du<sup>1</sup>; Zhuo Shao<sup>3</sup>; Maojun Ge<sup>4</sup>; Qingming Wang<sup>1</sup>; Wei Yang<sup>1,2</sup>  
11  
12

13 \*Contributed equally to this manuscript  
14  
15

- 16  
17 1. Department of Anorectal Surgery, Shuguang Hospital, Shanghai University of  
18 Traditional Chinese Medicine, Shanghai, China;  
19  
20 2. Anorectal Disease Institute of Shuguang Hospital, Shanghai, China;  
21  
22 3. Department of General Surgery, Changhai Hospital, The Second Military  
23 Medical University, Shanghai, China;  
24  
25 4. Department of General Surgery, Shuguang Hospital, Shanghai University of  
26 Traditional Chinese Medicine, Shanghai, China.  
27  
28  
29  
30

31 **Correspondence to:**

32 Dr. Zubing Mei, MD, PhD

33 Department of Anorectal Surgery, Shuguang Hospital, Shanghai University of  
34 Traditional Chinese Medicine; Anorectal Disease Institute of Shuguang Hospital  
35 528 Zhangheng Road, Shanghai 201203, China  
36  
37  
38

39 Tel: 086-2120256187

40 E-mail: herrmayor@126.com  
41  
42  
43  
44  
45  
46  
47

48 Professor. Wei Yang, MD

49 Department of Anorectal Surgery, Shuguang Hospital, Shanghai University of  
50 Traditional Chinese Medicine; Anorectal Disease Institute of Shuguang Hospital  
51 528 Zhangheng Road, Shanghai 201203, China  
52  
53  
54

55 Tel: 086-2120256187

56 E-mail: yangweiyishi@163.com  
57  
58  
59  
60

1  
2  
3  
4 **Running head:** Risk prediction models for AFR and related complications  
5

6 **Abstract Word Count:** 298  
7

8  
9 **Word Count (excluding abstract):** 3,257  
10

11 **References:** 53  
12

13  
14 **Figures:** 1  
15

16  
17  
18  
19 **Key words:** anal fistula, recurrence, prediction model, complication, surgery, cohort  
20  
21  
22 study  
23

24  
25  
26  
27 **Funding statement:** This work was supported by the National Natural Science  
28  
29 Foundation of China (grant no. 81774112), Science and Technology Commission of  
30  
31 Shanghai Municipality (grant no. 18401904600, 19401933900 and 16401971600),  
32  
33 TCM dominant disease breeding project (grant no. zyz-2017018), the Plateau Peak  
34  
35 Disciplines Innovation Team Project of Shanghai Municipal Education Commission  
36  
37  
38  
39  
40 (grant no. A1-N1920501010220).  
41  
42  
43  
44  
45

46 **Abbreviations:** AF, anal fistula; AFR, anal fistula recurrence; BMI, body mass index;  
47  
48 FCS, fully conditional specification; MVNI, multivariate normal imputation; TRIPOD,  
49  
50 Transparent Reporting of a multivariable prediction model for Individual Prognosis  
51  
52 Or Diagnosis.  
53  
54

55  
56  
57  
58  
59 **Author contributions:**  
60

1  
2  
3  
4 Dr. Zubing Mei had full access to all of the data in the study and takes responsibility  
5  
6 for the integrity of the data and the accuracy of the data analysis. Drs Mei and Li are  
7  
8 co-first authors of this article.  
9

10  
11 Study concept and design: Zubing Mei.  
12

13  
14 Acquisition, analysis, or interpretation of data: Yue Li, Zubing Mei, Zhijun Zhang, Ye  
15  
16 Han, Suzhi Liu, Haikun Zhou, Peixin Du, Qingming Wang, Wei Yang.  
17  
18

19  
20 Drafting of the manuscript: Zubing Mei.  
21

22  
23 Critical revision of the manuscript for important intellectual content: All authors.  
24

25  
26 Statistical analysis: Zubing Mei, Zhuo Shao, Maojun Ge.  
27

28  
29 Administrative, technical, or material support: All authors.  
30

31  
32 Study supervision: Zubing Mei.  
33  
34

35  
36 **Competing interests statement:**  
37

38  
39 The authors declare that they have no conflict of interest.  
40  
41

42  
43 **Role of the Funder/Sponsor:**  
44

45  
46 The funder of the study had no role in the study design, data collection, data analysis,  
47  
48 data interpretation or writing of the manuscript. The corresponding author had full  
49  
50 access to all the data in the study and has final responsibility for the decision to  
51  
52 submit for publication.  
53  
54  
55  
56  
57  
58  
59  
60

## **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

1  
2  
3  
4 prevention remains the best strategy. This study will provide alternative tools for early  
5  
6 screening of high-risk patients of AFR and related complications, helping surgeons  
7  
8 better understand the aetiology and outcome of anal fistula in an earlier stage.  
9  
10  
11  
12  
13

#### 14 **Ethics and dissemination**

15  
16  
17 The study is approved by the Institutional Review Board of Shuguang Hospital  
18  
19 affiliated with Shanghai University of TCM (Approval Number: 2019-699-54-01).  
20  
21

22 The results of this cohort will be submitted to international scientific peer-reviewed  
23  
24 journals or conferences in surgery, anorectal surgery or anorectal diseases.  
25  
26  
27  
28  
29

#### 30 **Trial registration number**

31  
32 Chinese Clinical Trial Registry (ChiCTR1900025069); Pre-results.  
33  
34  
35  
36  
37

#### 38 **Key words**

39  
40 Anal fistula; treatment outcome; recurrence; surgery; cohort study; prediction model  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60



### 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 questionnaire platform.
- Some more potentially predictors will not be involved or not collected in the current study, for example, data related to postoperative nursing strategy and outpatient follow-up frequency.

Review only

## **Introduction**

Anal fistula (AF) is a common perianal disease usually infected by cryptoglandular origin, which is regarded as a chronic stage of perianal abscess.<sup>1</sup> 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.<sup>2,3</sup> 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.<sup>4</sup> 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%.<sup>5-8</sup> 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.<sup>9,10</sup> 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.<sup>4,11-13</sup> 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.<sup>5</sup>

1  
2  
3  
4 Mei et al. conducted a meta-analysis involving 20 studies with 6168 patients and  
5  
6 concluded high transsphincteric fistula, internal opening unidentified, and horseshoe  
7  
8 extensions were independent risk factors for AFR with high-quality evidence, while  
9  
10 prior anal surgery, seton placement surgery, and multiple fistula tract were  
11  
12 demonstrated as risk factors for AFR with moderate-quality evidence.<sup>4</sup> Factors  
13  
14 influencing other perioperative complications related AF surgery including wound  
15  
16 hemorrhage, edema, urinary retention, delayed wound healing and unplanned  
17  
18 hospitalization are also rarely reported. Therefore, there is an urgent need to develop  
19  
20 risk prediction tools for the complete profile of risk factors for AFR and those related  
21  
22 complications.  
23  
24  
25  
26  
27  
28  
29  
30  
31

32 Which AF patients will be cured after surgery and which ones will not, are  
33  
34 rarely investigated. The development of a prediction model for AFR following  
35  
36 surgery to identify patients with high risk, would be of significant importance. Firstly,  
37  
38 surgeons can provide patients preoperatively about an estimated surgical cure rate  
39  
40 according to the prediction models. Moreover, the current knowledge in the literature  
41  
42 reporting potential predictive factors could instruct patients to avoid personal risk  
43  
44 factors and adjust treatment strategy in order to improve the surgical cure rate, which  
45  
46 have been well described and applied in the prevention of other diseases.<sup>14-17</sup>  
47  
48  
49  
50  
51  
52  
53  
54  
55

56 However, so far, there are no effective screening tools to evaluate and predict  
57  
58 the risk of recurrence or other adverse outcomes of AF. Therefore, the aim of the  
59  
60

1  
2  
3  
4 current study was to develop and validate multivariable prediction models that predict  
5  
6 postoperative AFR and related complications.  
7  
8  
9

### 10 11 **Aims and objectives** 12

13  
14 The aim of this study is to develop risk prediction models for postoperative  
15  
16 recurrence as well as other surgery-related complications in a prospective  
17  
18 hospital-based AF cohort. Risk prediction model for perioperative complications  
19  
20 including wound hemorrhage, edema, urinary retention, delayed wound healing and  
21  
22 unplanned hospitalization will also be developed. Flowchart of prediction model  
23  
24 development and assessment is provided in Figure 1.  
25  
26  
27  
28

29  
30 The detailed tasks of this study are to:  
31

- 32 1. Calculate the 3 to 6-month incidence of recurrence, and related  
33  
34 complications in patients following AF surgery.  
35  
36
- 37 2. Establish the risk factors that significantly predict postoperative AFR and  
38  
39 related complications of an AF cohort in a tertiary referral center.  
40  
41
- 42 3. Develop and validate risk prediction models for postoperative AFR and  
43  
44 related complications of an AF cohort in a tertiary referral center.  
45  
46  
47

48 We also have the following two hypotheses examined:  
49

- 50 1. Patient-related demographic characteristics, fistula and surgery-related  
51  
52 factors are predictive of postoperative AFR and related complications as dependent  
53  
54 variables.  
55  
56  
57  
58  
59  
60

1  
2  
3  
4 2. The risk prediction models for postoperative AFR and related complications  
5  
6 developed in our study have more than 70% of discriminating power.  
7  
8  
9

## 10 11 **Patients and Methods**

### 12 13 **Study design and participants**

14  
15 This study is a single-center, prospective observational study on a hospital-based  
16  
17 cohort enrolled at a tertiary referral center in Shanghai, China.  
18  
19  
20  
21  
22  
23  
24

### 25 26 **Eligibility criteria**

27  
28 The enrollment of the cohort subject was initiated from June, 2019. All subjects who  
29  
30 will undergo surgical intervention for AF will be included for inclusion. All  
31  
32 operations will be performed by a group of colon and rectal surgeons at Shuguang  
33  
34 Hospital, a regional tertiary referral center. Exclusion criteria are those whose age <  
35  
36 18 years, non-cryptoglandular fistula (eg, anal fistula due to inflammatory bowel  
37  
38 disease, human immunodeficiency virus, malignant cancer, or obstetrical trauma), and  
39  
40 rectovaginal or rectourethral fistula. The electronic medical records of the included  
41  
42 subjects should be complete.  
43  
44  
45  
46  
47  
48  
49  
50

51 Trained clinical investigators are collecting data in several categories, including  
52  
53 baseline demographics, laboratory examinations, surgical profiles, colonoscopic and  
54  
55 MR imaging findings and postoperative outcomes within 3 to 6 months. Planned  
56  
57  
58  
59  
60

1  
2  
3  
4 clinical reviews or electronic surveys are conducted during hospitalization and every  
5  
6 half to 3 months after discharge for 6 months.  
7  
8  
9

### 10 11 **Data collection**

12  
13  
14 The research team are all anorectal surgeons consisting of a principal investigator and  
15  
16 supervised by the Ethics committee of Shuguang Hospital. Written informed consent  
17  
18 was obtained from all patients. Investigators will not intervene in any aspects of  
19  
20 patient surveys at every stage of follow-up. Data are collected using a convenient  
21  
22 follow-up system supported by Empower EDC (OpenClinica, Boston, Massachusetts,  
23  
24 USA). This electronic system introduces a machine learning algorithm, through which  
25  
26 we can use the data already entered in the Empower system to train the algorithm  
27  
28 model and let the system itself develop quality control algorithms, validate the entered  
29  
30 data and identify missing or suspicious data. Finally, the data manager will check the  
31  
32 missing or suspicious data, confirm their completeness and asked the data manager to  
33  
34 provide additional data when necessary. Furthermore, an automatic reminder  
35  
36 follow-up function also plays a pivotal role during the whole follow-up period.  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47

### 48 **Patient and public involvement**

49  
50 Patients and public will not be involved in the development, design, conduct or  
51  
52 reporting of the study. The general results will be disseminated to participants through  
53  
54 public education during follow-up.  
55  
56  
57  
58  
59  
60

## 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.<sup>23 18</sup> 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<sup>4</sup> 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<sup>4</sup> 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  $p < 0.05$ .

1  
2  
3  
4  
5  
6  
7 Factors identified from our recent systematic review and Delphi survey (manuscript  
8  
9 under review), will be measured at baseline. These include factors involving the  
10  
11 identified significant risk factors which are reported in our meta-analysis are  
12  
13 presented as follows:  
14

- 15  
16 ▶ Prior anal surgery.
- 17  
18 ▶ Seton placement surgery.
- 19  
20 ▶ High transsphincteric fistula.
- 21  
22 ▶ Internal opening unidentified.
- 23  
24 ▶ Horseshoe extensions.
- 25  
26 ▶ Multiple fistula tracts.
- 27  
28  
29  
30  
31  
32  
33

34  
35 Some of the demographic factors and surgery details will also be collected due to  
36  
37 limited power in our literature review and some non-significant potential factors (eg,  
38  
39 smoking or alcohol use) may be risk factors and are also included as follows:  
40  
41

- 42  
43 ▶ Gender.
- 44  
45 ▶ Age.
- 46  
47 ▶ Smoking.
- 48  
49 ▶ Alcohol use.
- 50  
51 ▶ Diabetes mellitus.
- 52  
53 ▶ Obesity.
- 54  
55 ▶ Preoperative seton drainage.
- 56  
57  
58  
59  
60



1  
2  
3  
4 ▶ High internal opening.

5  
6 ▶ Postoperative drainage.

7  
8  
9 ▶ Suprlevator extensions.

10  
11 Other factors like laboratory examinations and MR imaging parameters (height of the  
12 internal openings, height and number of fistula, etc.) will also be collected.  
13  
14  
15

### 16 17 18 19 **Categorization of Potential Predictors**

20  
21 For categorical predictors, we can code them as “factor” variables, with coding as  
22 dummy variables, for example, smoking is coded originally as “1” for never smoker,  
23 “2” for past smoker, and “3” for current smoker and never smoker was selected as the  
24 reference category. Similar manner can be applied with alcohol use.<sup>19</sup>  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34

35 Continuous variables formally should be measured at an interval or ratio scale, and  
36 should be able to take any value in a range. We treat ordered variables as linear which  
37 is generally reasonable for prediction. In other cases, continuous predictors can be  
38 grouped with meaningful categorization; for example, body mass index (BMI) can be  
39 classified based on internationally recognized categories (i.e., underweight, normal  
40 weight, overweight, and obesity).<sup>20</sup> Based on previous experiences, we will be  
41 deriving some predictors based on the responses of the surveys. However, in case  
42 some subjectivity in the classifications of these predictors may occur, sensitivity  
43 analyses will be performed to examine the robustness of our definitions during model  
44 building and validation.  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

## 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.<sup>21 22</sup> 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,<sup>23-26</sup> which can reasonably approximate the true distributional relationship between the missing values and the available ones.<sup>27</sup> 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.<sup>27</sup>

## 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 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.<sup>28 29</sup> 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  $\geq 50\%$  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.<sup>30</sup> 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.<sup>23 24 26</sup>

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  $>10$ .<sup>31</sup> Then we will examine the interactions among the regression models.

## 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.<sup>32 33</sup> 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).<sup>34</sup> 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.<sup>35</sup> 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.<sup>34</sup> 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.<sup>36</sup> Calibration-in-the-large, which defines as the agreement between mean observed outcomes and mean predictions, will also be assessed for calibration.<sup>37</sup>

### **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.<sup>34 38</sup> 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.<sup>39</sup> We will also apply temporal validation as external validation using a more recent AF patient cohort.<sup>40</sup>

### **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.<sup>41-43</sup> 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.<sup>4 5</sup> 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).

### **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.<sup>44</sup> 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,<sup>45-47</sup> lifestyle

1  
2  
3  
4 factors such as smoking, alcohol abuse,<sup>48 49</sup> fistula-related factors such as number of  
5  
6 fistula tracts, horseshoe extensions, classification and location of fistula,  
7  
8  
9 surgery-related variables such as prior anal surgery and postoperative drainage<sup>48 50 51</sup>)  
10  
11 easily ascertainable before surgery may predict AFR. Similar risk prediction models  
12  
13 exist in other diseases, such as the Framingham Risk Score model to predict  
14  
15 cardiovascular disease risk<sup>52</sup> and the Korean Crohn's Disease Prediction (KCDP)  
16  
17 model to predict the clinical course of Crohn's disease.<sup>53</sup> Until now, risk factor  
18  
19 investigation of predictors of perioperative surgery-related complications has been  
20  
21 limited to assessment of single predictors with small sample size. The risk prediction  
22  
23 models can help inform surgeons regarding high risk AF patients based on the overall  
24  
25 risk factors. The primary purpose of study was to develop two risk prediction models  
26  
27 to facilitate surgeons in identifying AF patients who will have surgical treatment at  
28  
29 higher risk of developing recurrence and surgery-related complications. The  
30  
31 predictive models will help both clinicians and patients identify the risk of  
32  
33 complications after AF surgery in advance, take necessary interventions to reduce  
34  
35 the risk of surgery-related complications as well as the personal and social financial  
36  
37 burden brought about by those complications. The accurate risk prediction models are  
38  
39 especially instructive for the development of the optimal surgical plan to achieve  
40  
41 optimal surgical outcome.

42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56 Our study has several strengths. To the best of our knowledge, this is the first study  
57  
58 with the primary aim to develop, internally and externally validate multivariable  
59  
60

1  
2  
3  
4 prediction models for AFR and related complications following AF surgery.  
5

6 Multidimensional clinically useful candidate predictors will be fully examined from a  
7  
8 variety of sources including our published systematic review, Delphi surveys and  
9  
10 univariable or multivariable logistic regression analysis. Second, we will apply the  
11  
12 internal and external validation of the prediction models using bootstrapping  
13  
14 procedure. Third, multiple imputation will also be used to treat the missing data. Last  
15  
16 but not least, our study is a prospective cohort one with adequate follow-up period  
17  
18 which can minimize certain bias.  
19  
20  
21  
22  
23  
24  
25  
26

27 Our study also has limitations. As many of the variables are collected through a  
28  
29 Wechat questioner platform during the hospitalization and follow-up, non-response  
30  
31 bias can occur. To solve this issue, we regularly send reminders to those who does not  
32  
33 respond after discharge. Second, though we will investigate a series of potential  
34  
35 predictors, some more potentially predictors will not be involved or not collected in  
36  
37 the current study, data related to postoperative nursing strategy and outpatient  
38  
39 follow-up frequency for example.  
40  
41  
42  
43  
44  
45  
46  
47

48 The newly developed risk algorithms may have significant applications in clinical  
49  
50 practice by helping recommend optimal surgical approach for a specific AF patient,  
51  
52 as well as intensive perioperative care and education, timely assessment and  
53  
54 discussion of the need for interventions to those most at high risk of developing  
55  
56 recurrence or surgery-related complications. The models will specifically identify AF  
57  
58  
59  
60



1  
2  
3  
4 patients who are likely to develop recurrence or related complications following AF  
5  
6 surgery to offer the quantitative evaluation of the risk. Moreover, the models will also  
7  
8 provide reference information for preventing recurrence and reducing the rate of  
9  
10 recurrence after operation, and to intervene some high risk factors in the early stage.  
11  
12  
13  
14  
15  
16

## 17 **CONCLUSION**

18  
19 This study protocol summarizes the design of development and validation studies for  
20  
21 a risk screening tool in patients receiving AF surgery. Results from this study will be  
22  
23 interpreted for the purpose of clinical decision making. The models to be developed of  
24  
25 the study could be used to make new recommendations for perioperative AF patients.  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

## References

1. Tantiplachiva 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.ijvsu.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

- 1  
2  
3  
4 7. Gottgens KW, Janssen PT, Heemskerk J, et al. Long-term outcome of low perianal  
5  
6 fistulas treated by fistulotomy: a multicenter study. *International journal of*  
7  
8 *colorectal disease* 2015;30(2):213-9. doi: 10.1007/s00384-014-2072-y  
9  
10  
11  
12 8. Bakhtawar N, Usman M. Factors Increasing the Risk of Recurrence in Fistula-in-ano.  
13  
14 *Cureus* 2019;11(3):e4200. doi: 10.7759/cureus.4200  
15  
16  
17  
18 9. Sygut A, Mik M, Trzcinski R, et al. How the location of the internal opening of anal  
19  
20 fistulas affect the treatment results of primary transsphincteric fistulas.  
21  
22 *Langenbeck's archives of surgery* 2010;395(8):1055-9. doi:  
23  
24 10.1007/s00423-009-0562-0  
25  
26  
27  
28  
29  
30 10. Dziki A, Bartos M. Seton treatment of anal fistula: experience with a new  
31  
32 modification. *The European journal of surgery = Acta chirurgica* 1998;164(7):543-8.  
33  
34 doi: 10.1080/110241598750005930  
35  
36  
37  
38  
39 11. Hermann J, Eder P, Banasiewicz T, et al. Current management of anal fistulas in  
40  
41 Crohn's disease. *Przegląd gastroenterologiczny* 2015;10(2):83-8. doi:  
42  
43 10.5114/pg.2015.49684  
44  
45  
46  
47  
48 12. Hamadani A, Haigh PI, Liu IL, et al. Who is at risk for developing chronic anal  
49  
50 fistula or recurrent anal sepsis after initial perianal abscess? *Diseases of the colon*  
51  
52 *and rectum* 2009;52(2):217-21. doi: 10.1007/DCR.0b013e31819a5c52  
53  
54  
55  
56  
57  
58  
59  
60

- 1  
2  
3  
4 13. Ellis CN, Clark S. Effect of tobacco smoking on advancement flap repair of  
5  
6 complex anal fistulas. *Diseases of the colon and rectum* 2007;50(4):459-63. doi:  
7  
8 10.1007/s10350-006-0829-2  
9  
10  
11  
12 14. Moons KG, Altman DG, Vergouwe Y, et al. Prognosis and prognostic research:  
13  
14 application and impact of prognostic models in clinical practice. *Bmj* 2009;338:b606.  
15  
16 doi: 10.1136/bmj.b606  
17  
18  
19  
20  
21 15. Ming C, Viassolo V, Probst-Hensch N, et al. Machine learning techniques for  
22  
23 personalized breast cancer risk prediction: comparison with the BCRAT and  
24  
25 BOADICEA models. *Breast cancer research : BCR* 2019;21(1):75. doi:  
26  
27 10.1186/s13058-019-1158-4  
28  
29  
30  
31  
32  
33 16. Khan SS, Ning H, Shah SJ, et al. 10-Year Risk Equations for Incident Heart Failure  
34  
35 in the General Population. *Journal of the American College of Cardiology*  
36  
37 2019;73(19):2388-97. doi: 10.1016/j.jacc.2019.02.057  
38  
39  
40  
41  
42 17. Terry MB, Liao Y, Whittemore AS, et al. 10-year performance of four models of  
43  
44 breast cancer risk: a validation study. *The Lancet Oncology* 2019;20(4):504-17. doi:  
45  
46 10.1016/S1470-2045(18)30902-1  
47  
48  
49  
50  
51 18. Yano T, Asano M, Matsuda Y, et al. Prognostic factors for recurrence following  
52  
53 the initial drainage of an anorectal abscess. *International journal of colorectal*  
54  
55 *disease* 2010;25(12):1495-8. doi: 10.1007/s00384-010-1011-9 [published Online  
56  
57 First: 2010/07/20]  
58  
59  
60

1  
2  
3  
4 19. Steyerberg EW. Coding of Categorical and Continuous Predictors. In: Steyerberg  
5  
6 EW, ed. Clinical Prediction Models: A Practical Approach to Development, Validation,  
7  
8 and Updating. Cham: Springer International Publishing 2019:175-90.  
9

10  
11  
12 20. Morgan TM, Elashoff RM. Effect of Categorizing a Continuous Covariate on the  
13  
14 Comparison of Survival Time. Journal of the American Statistical Association  
15  
16 1986;81(396):917-21. doi: 10.1080/01621459.1986.10478350  
17  
18

19  
20  
21 21. van Oort L, van den Berg T, Koes BW, et al. Preliminary state of development of  
22  
23 prediction models for primary care physical therapy: a systematic review. Journal of  
24  
25 Clinical Epidemiology 2012;65(12):1257-66. doi:  
26  
27 <https://doi.org/10.1016/j.jclinepi.2012.05.007>  
28  
29

30  
31  
32 22. Rushton AB, Verra ML, Emms A, et al. Development and validation of two clinical  
33  
34 prediction models to inform clinical decision-making for lumbar spinal fusion surgery  
35  
36 for degenerative disorders and rehabilitation following surgery: protocol for a  
37  
38 prospective observational study. BMJ Open 2018;8(5):e021078. doi:  
39  
40 [10.1136/bmjopen-2017-021078](https://doi.org/10.1136/bmjopen-2017-021078)  
41  
42

43  
44  
45 23. Wood AM, White IR, Royston P. How should variable selection be performed  
46  
47 with multiply imputed data? Statistics in Medicine 2008;27(17):3227-46. doi:  
48  
49 [10.1002/sim.3177](https://doi.org/10.1002/sim.3177)  
50  
51

- 1  
2  
3  
4 24. White IR, Royston P, Wood AM. Multiple imputation using chained equations:  
5  
6 Issues and guidance for practice. *Statistics in Medicine* 2011;30(4):377-99. doi:  
7  
8 10.1002/sim.4067  
9  
10  
11  
12 25. Schafer JL. Multiple imputation: a primer. *Statistical Methods in Medical*  
13  
14 *Research* 1999;8(1):3-15. doi: 10.1177/096228029900800102  
15  
16  
17  
18 26. Marshall A, Altman DG, Holder RL, et al. Combining estimates of interest in  
19  
20 prognostic modelling studies after multiple imputation: current practice and  
21  
22 guidelines. *BMC Medical Research Methodology* 2009;9(1):57. doi:  
23  
24 10.1186/1471-2288-9-57  
25  
26  
27  
28  
29 27. Lee KJ, Carlin JB. Multiple Imputation for Missing Data: Fully Conditional  
30  
31 Specification Versus Multivariate Normal Imputation. *American Journal of*  
32  
33 *Epidemiology* 2010;171(5):624-32. doi: 10.1093/aje/kwp425  
34  
35  
36  
37  
38 28. Sauerbrei W, Royston P, Binder H. Selection of important variables and  
39  
40 determination of functional form for continuous predictors in multivariable model  
41  
42 building. *Statistics in Medicine* 2007;26(30):5512-28. doi: 10.1002/sim.3148  
43  
44  
45  
46  
47 29. Royston P, Sauerbrei W. Building multivariable regression models with  
48  
49 continuous covariates in clinical epidemiology--with an emphasis on fractional  
50  
51 polynomials. *Methods of information in medicine* 2005;44(4):561-71.  
52  
53  
54  
55  
56 30. Mantel N. Why Stepdown Procedures in Variable Selection. *Technometrics*  
57  
58 1970;12(3):621-25. doi: 10.1080/00401706.1970.10488701  
59  
60

1  
2  
3  
4 31. Vittinghoff E, Glidden DV, Shiboski SC, et al. Regression methods in biostatistics:  
5  
6 linear, logistic, survival, and repeated measures models: Springer Science & Business  
7  
8 Media 2011.  
9

10  
11  
12 32. Steyerberg EW, Vickers AJ, Cook NR, et al. Assessing the performance of  
13  
14 prediction models: a framework for traditional and novel measures. *Epidemiology*  
15  
16 2010;21(1):128-38. doi: 10.1097/EDE.0b013e3181c30fb2  
17  
18  
19

20  
21 33. NAGELKERKE NJD. A note on a general definition of the coefficient of  
22  
23 determination. *Biometrika* 1991;78(3):691-92. doi: 10.1093/biomet/78.3.691  
24  
25  
26

27  
28 34. Harrell Jr FE. Regression modeling strategies: with applications to linear models,  
29  
30 logistic and ordinal regression, and survival analysis: Springer 2015.  
31  
32

33  
34 35. Rubin DB. Multiple imputation for nonresponse in surveys: John Wiley & Sons  
35  
36 2004.  
37  
38

39  
40 36. Hosmer Jr DW, Lemeshow S, Sturdivant RX. Applied logistic regression: John  
41  
42 Wiley & Sons 2013.  
43  
44

45  
46 37. Snell KI, Ensor J, Debray TP, et al. Meta-analysis of prediction model performance  
47  
48 across multiple studies: Which scale helps ensure between-study normality for the  
49  
50 C-statistic and calibration measures? *Statistical methods in medical research*  
51  
52 2018;27(11):3505-22.  
53  
54  
55  
56  
57  
58  
59  
60

1  
2  
3  
4 38. Chatfield C. Model uncertainty, data mining and statistical inference. *Journal of*  
5  
6  
7 the Royal Statistical Society: Series A (Statistics in Society) 1995;158(3):419-44.

8  
9  
10 39. Steyerberg EW, Bleeker SE, Moll HA, et al. Internal and external validation of  
11  
12  
13 predictive models: a simulation study of bias and precision in small samples. *Journal*  
14  
15  
16 of clinical epidemiology 2003;56(5):441-47.

17  
18  
19 40. Justice AC, Covinsky KE, Berlin JA. Assessing the generalizability of prognostic  
20  
21  
22 information. *Annals of internal medicine* 1999;130(6):515-24.

23  
24  
25 41. Concato J, Peduzzi P, Holford TR, et al. Importance of events per independent  
26  
27  
28 variable in proportional hazards analysis I. Background, goals, and general strategy.  
29  
30  
31 *Journal of clinical epidemiology* 1995;48(12):1495-501.

32  
33  
34 42. Peduzzi P, Concato J, Kemper E, et al. A simulation study of the number of events  
35  
36  
37 per variable in logistic regression analysis. *Journal of clinical epidemiology*  
38  
39  
40 1996;49(12):1373-79.

41  
42  
43 43. Steyerberg EW. *Clinical prediction models: A practical approach to development,*  
44  
45  
46 *validation, and updating* Springer 2009.

47  
48  
49 44. Moons KG, Altman DG, Reitsma JB, et al. Transparent Reporting of a  
50  
51  
52 multivariable prediction model for Individual Prognosis or Diagnosis (TRIPOD):  
53  
54  
55 explanation and elaboration. *Annals of internal medicine* 2015;162(1):W1-W73.



- 1  
2  
3  
4 45. Hamadani A, Haigh PI, Liu I-LA, et al. Who is at risk for developing chronic anal  
5  
6 fistula or recurrent anal sepsis after initial perianal abscess? *Diseases of the colon &*  
7  
8 *rectum* 2009;52(2):217-21.  
9  
10  
11  
12 46. Sonoda T, Hull T, Piedmonte MR, et al. Outcomes of primary repair of anorectal  
13  
14 and rectovaginal fistulas using the endorectal advancement flap. *Diseases of the*  
15  
16 *colon & rectum* 2002;45(12):1622-28.  
17  
18  
19  
20 47. Hyman N, O'Brien S, Osler T. Outcomes after fistulotomy: results of a prospective,  
21  
22 multicenter regional study. *Diseases of the Colon & Rectum* 2009;52(12):2022-27.  
23  
24  
25  
26 48. Ellis CN, Clark S. Effect of tobacco smoking on advancement flap repair of  
27  
28 complex anal fistulas. *Diseases of the colon & rectum* 2007;50(4):459-63.  
29  
30  
31  
32 49. Schwandner T, Roblick MH, Kierer W, et al. Surgical treatment of complex anal  
33  
34 fistulas with the anal fistula plug: a prospective, multicenter study. *Diseases of the*  
35  
36 *Colon & Rectum* 2009;52(9):1578-83.  
37  
38  
39  
40  
41  
42 50. Garcia-Aguilar J, Belmonte C, Wong WD, et al. Anal fistula surgery: Factors  
43  
44 associated with recurrence and incontinence. *Diseases of the colon & rectum*  
45  
46 1996;39(7):723-29.  
47  
48  
49  
50  
51 51. Athanasiadis S, Köhler A, Nafe M. Treatment of high anal fistulae by primary  
52  
53 occlusion of the internal ostium, drainage of the intersphincteric space, and mucosal  
54  
55 advancement flap. *International journal of colorectal disease* 1994;9(3):153-57.  
56  
57  
58  
59  
60

1  
2  
3  
4 52. Fox ER, Samdarshi TE, Musani SK, et al. Development and validation of risk  
5  
6 prediction models for cardiovascular events in black adults: the Jackson Heart Study  
7  
8 Cohort. JAMA cardiology 2016;1(1):15-25.  
9  
10

11  
12 53. Park Y, Cheon JH, Park YL, et al. Development of a novel predictive model for the  
13  
14 clinical course of Crohn's disease: results from the connect study. Inflammatory  
15  
16 bowel diseases 2017;23(7):1071-79.  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

**Figure legend**

Figure 1. Flowchart of prediction model development and assessment.

For peer review only

Figure 1: Flowchart of prediction model development and assessment.

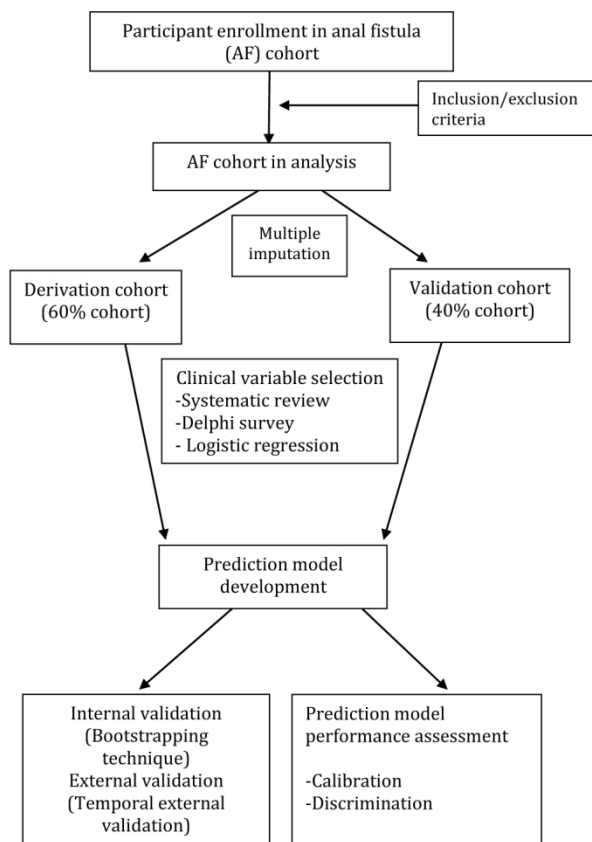


Figure 1. Flowchart of prediction model development and assessment.

## 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
<b>Title and abstract</b>				
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
<b>Introduction</b>				
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
<b>Methods</b>				
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
<b>Results</b>				
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	-



TRIPOD Checklist: Prediction Model Development and Validation

specification	15b	D	coefficients, and model intercept or baseline survival at a given time point). 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).	-
<b>Discussion</b>				
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
<b>Other information</b>				
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.

## 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
<b>Title and abstract</b>				
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
<b>Introduction</b>				
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
<b>Methods</b>				
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
<b>Results</b>				
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	-



## TRIPOD Checklist: Prediction Model Development and Validation

specification	15b	D	coefficients, and model intercept or baseline survival at a given time point). 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).	-
<b>Discussion</b>				
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
<b>Other information</b>				
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.



# 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:	<i>BMJ Open</i>
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
<b>Primary Subject Heading</b>:	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



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.

1  
2  
3  
4 **Development of screening tools to predict the risk of recurrence and related**  
5 **complications following anal fistula surgery: protocol for a prospective cohort**  
6 **study**  
7  
8  
9

10  
11 Zubing Mei<sup>1,2\*</sup>; Yue Li<sup>1\*</sup>; Zhijun Zhang<sup>1</sup>; Haikun Zhou<sup>1</sup>; Suzhi Liu<sup>1</sup>; Ye Han<sup>1</sup>;  
12 Peixin Du<sup>1</sup>; Xiufang Qin<sup>3</sup>; Zhuo Shao<sup>4</sup>; Maojun Ge<sup>5</sup>; Qingming Wang<sup>1</sup>; Wei Yang<sup>1,2</sup>

13  
14  
15 \*Contributed equally to this manuscript  
16

- 17  
18  
19 1. Department of Anorectal Surgery, Shuguang Hospital, Shanghai University of  
20 Traditional Chinese Medicine, Shanghai, China;  
21  
22 2. Anorectal Disease Institute of Shuguang Hospital, Shanghai, China;  
23  
24 3. Department of Nursing, Shuguang Hospital, Shanghai University of  
25 Traditional Chinese Medicine, Shanghai, China;  
26  
27 4. Department of General Surgery, Changhai Hospital, The Second Military  
28 Medical University, Shanghai, China;  
29  
30 5. Department of General Surgery, Shuguang Hospital, Shanghai University of  
31 Traditional Chinese Medicine, Shanghai, China.  
32  
33  
34  
35

36  
37  
38 **Correspondence to:**

39 Dr. Zubing Mei, MD, PhD

40  
41 Department of Anorectal Surgery, Shuguang Hospital, Shanghai University of  
42 Traditional Chinese Medicine; Anorectal Disease Institute of Shuguang Hospital  
43 528 Zhangheng Road, Shanghai 201203, China  
44  
45

46 Tel: 086-2120256187

47  
48 E-mail: herrmayor@126.com  
49

50  
51  
52  
53 Professor. Wei Yang, MD

54 Department of Anorectal Surgery, Shuguang Hospital, Shanghai University of  
55 Traditional Chinese Medicine; Anorectal Disease Institute of Shuguang Hospital  
56 528 Zhangheng Road, Shanghai 201203, China  
57  
58  
59  
60

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. zyz-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. shslczdk04302), 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).

1  
2  
3  
4  
5  
6 **Abbreviations:** AF, anal fistula; AFR, anal fistula recurrence; BMI, body mass index;  
7  
8  
9 FCS, fully conditional specification; MVNI, multivariate normal imputation; TRIPOD,  
10  
11 Transparent Reporting of a multivariable prediction model for Individual Prognosis  
12  
13 Or Diagnosis.  
14  
15

16  
17  
18  
19 **Author contributions:**

20  
21  
22 Dr. Zubing Mei had full access to all of the data in the study and takes responsibility  
23  
24 for the integrity of the data and the accuracy of the data analysis. Drs Mei and Li are  
25  
26 co-first authors of this article.  
27  
28

29  
30 Study concept and design: Zubing Mei.

31  
32 Acquisition, analysis, or interpretation of data: Yue Li, Zubing Mei, Zhijun Zhang, Ye  
33  
34 Han, Suzhi Liu, Haikun Zhou, Peixin Du, Xiufang Qin, Qingming Wang, Wei Yang.

35  
36 Drafting of the manuscript: Zubing Mei.

37  
38 Critical revision of the manuscript for important intellectual content: All authors.

39  
40 Statistical analysis: Zubing Mei, Zhuo Shao, Maojun Ge.

41  
42 Administrative, technical, or material support: All authors.

43  
44 Study supervision: Zubing Mei.  
45  
46  
47

48  
49  
50  
51  
52 **Competing interests statement:**

53  
54  
55 The authors declare that they have no conflict of interest.  
56  
57  
58  
59  
60

**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 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

1  
2  
3  
4 warn, motivate and empower patients to avoid some modifiable risk factors to early  
5  
6 prevent postoperative complications. This study will also provide alternative tools for  
7  
8 early screening of high-risk patients of AFR and related complications, helping  
9  
10 surgeons better understand the aetiology and outcome of AF in an earlier stage.  
11  
12  
13  
14  
15  
16

### 17 **Ethics and dissemination**

18  
19 The study is approved by the Institutional Review Board of Shuguang Hospital  
20  
21 affiliated with Shanghai University of Traditional Chinese Medicine (Approval  
22  
23 Number: 2019-699-54-01). The results of this study will be submitted to international  
24  
25 scientific peer-reviewed journals or conferences in surgery, anorectal surgery or  
26  
27 anorectal diseases.  
28  
29  
30  
31  
32  
33  
34

### 35 **Trial registration number**

36  
37 Chinese Clinical Trial Registry (ChiCTR1900025069); Pre-results.  
38  
39  
40  
41  
42

### 43 **Key words**

44  
45 Anal fistula; treatment outcome; recurrence; surgery; cohort study; prediction model  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60



### 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 ( $\geq 20$ ) 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 questionnaire system.

## **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.<sup>1</sup> 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,<sup>2,3</sup> 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.<sup>4</sup> 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%.<sup>5-8</sup> 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.<sup>9,10</sup> 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.<sup>4,11-13</sup> 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

1  
2  
3  
4 independent risk factors for AFR.<sup>5</sup> Recently, according to the evidence grading  
5  
6 criteria based on Egger's P value, total sample size and between-study heterogeneity,  
7  
8 we published a meta-analysis involving 20 studies with 6168 patients and concluded  
9  
10 high transsphincteric fistula, internal opening unidentified, and horseshoe extensions  
11  
12 were independent risk factors for AFR with high-quality evidence, while prior anal  
13  
14 surgery, seton placement surgery, and multiple fistula tract were demonstrated as risk  
15  
16 factors for AFR with moderate-quality evidence.<sup>4</sup>  
17  
18  
19  
20  
21  
22  
23  
24

25 Factors influencing other perioperative complications related AF surgery  
26  
27 including wound hemorrhage, fecal impaction, urinary retention, delayed wound  
28  
29 healing and unplanned hospitalization are also rarely reported. Therefore, there is an  
30  
31 urgent need to develop risk prediction tools for the complete profile of risk factors for  
32  
33 AFR and those related complications.  
34  
35  
36  
37  
38  
39

40 Which AF patients will be cured after surgery and which ones will not, are  
41  
42 rarely investigated. The development of a prediction model for AFR following  
43  
44 surgery to identify those patients with a higher risk of developing complications  
45  
46 during follow up, would be of significant importance. Firstly, surgeons can provide  
47  
48 patients preoperatively about an estimated surgical cure rate according to the  
49  
50 prediction models. Moreover, the current knowledge in the literature reporting  
51  
52 potential predictive factors could help patients know well their individual risk factors  
53  
54  
55  
56  
57  
58  
59  
60

1  
2  
3  
4 and avoid modifiable ones in order to improve the cure rate, which have been well  
5  
6 described and applied in the prevention of other diseases.<sup>14-17</sup>  
7  
8  
9

10  
11 However, so far, there are no effective screening tools to evaluate and predict  
12  
13 the risk of recurrence or other adverse outcomes of AF. Therefore, the aim of the  
14  
15 current study was to develop and validate multivariable prediction models that predict  
16  
17 postoperative AFR and related complications.  
18  
19  
20  
21  
22  
23  
24

### 25 **Aims and objectives**

26  
27 The aim of this study is to develop risk prediction models for postoperative  
28  
29 recurrence as well as other surgery-related complications in a prospective  
30  
31 hospital-based AF cohort. Risk prediction model for perioperative complications will  
32  
33 also be developed. Flowchart of prediction model development and assessment is  
34  
35 provided in Figure 1.  
36  
37  
38  
39

40 The detailed tasks of this study are to:

- 41  
42 1. Calculate the 3 to 6-month incidence of recurrence, and related  
43  
44 complications in patients following AF surgery.  
45  
46
- 47  
48 2. Establish the risk factors that significantly predict postoperative AFR and  
49  
50 related complications based on the AF cohort in a tertiary referral center.  
51  
52
- 53  
54 3. Develop and validate the risk prediction models for postoperative AFR and  
55  
56 related complications.  
57  
58  
59  
60

1  
2  
3  
4 4. Considering the different scenario for different surgical interventions,  
5  
6 stratified analyses are conducted based on surgery type. If possible, risk prediction  
7  
8 model will also be developed in relevant sub-populations, such as those only  
9  
10 receiving fistulectomy or fistulotomy, which can account for more than 60% of our  
11  
12 AF cohort populations.  
13  
14

15  
16  
17 We also have the following two hypotheses examined:  
18

19  
20 1. Patient-related demographic characteristics, fistula and surgery-related  
21  
22 factors are predictive of postoperative AFR and related complications as dependent  
23  
24 variables.  
25

26  
27 2. The risk prediction models for postoperative AFR and related complications  
28  
29 developed in our study have more than 70% of discriminating power.  
30  
31  
32  
33  
34

## 35 **Patients and Methods**

### 36 37 **Study design and participants**

38  
39  
40 This study is a single-center, prospective observational study on a hospital-based  
41  
42 cohort enrolled at a tertiary referral center in Shanghai, China.  
43  
44  
45  
46  
47

### 48 **Eligibility criteria**

49  
50 The enrollment of the cohort subject was initiated from June, 2019. All subjects who  
51  
52 will undergo surgical intervention for AF will be included for inclusion. All  
53  
54 operations will be performed by a group of colon and rectal surgeons at Shuguang  
55  
56 Hospital, a regional tertiary referral center. Exclusion criteria are those whose age <  
57  
58  
59  
60

1  
2  
3  
4 18 years, non-cryptoglandular fistula (eg, anal fistula due to inflammatory bowel  
5  
6 disease, human immunodeficiency virus, malignant cancer, or obstetrical trauma), and  
7  
8  
9 rectovaginal or rectourethral fistula. The electronic medical records of the included  
10  
11 subjects should be complete.  
12  
13  
14  
15  
16

17 Trained clinical investigators are collecting data in several categories, including  
18  
19 baseline demographics, laboratory examinations, clinical data, imaging findings and  
20  
21 follow-up information 3 to 6 months postoperatively. Planned clinical reviews or  
22  
23 electronic surveys are conducted during hospitalization and every half to 3 months  
24  
25 after discharge for 6 months.  
26  
27  
28  
29  
30  
31

### 32 **Data collection**

33  
34  
35 The research team comprised a principal investigator, 5 to 8 anorectal surgeons who  
36  
37 are trained and supervised by the Ethics committee of Shuguang Hospital. Written  
38  
39 informed consent was obtained from all patients. Investigators will not intervene in  
40  
41 any aspects of patient surveys at every stage of data collection and follow-up. Data  
42  
43 are collected using a convenient follow-up system supported by Empower EDC  
44  
45 (Solutions, Boston, Massachusetts, USA). This electronic system introduces a  
46  
47 machine learning algorithm, through which we can use the data already entered in the  
48  
49 Empower system to train the algorithm model and let the system itself develop quality  
50  
51 control algorithms, validate the entered data and identify missing or suspicious data.  
52  
53  
54  
55  
56  
57  
58 Finally, the data manager will check the missing or suspicious data, confirm their  
59  
60

1  
2  
3  
4 completeness and asked the data manager to provide additional data when necessary.  
5

6 Furthermore, an automatic reminder follow-up function also plays a pivotal role  
7  
8 during the whole follow-up period.  
9  
10

### 11 12 13 14 **Patient and public involvement**

15  
16 Patients and public will not be involved in the development, design, conduct or  
17  
18 reporting of the study. The general results will be disseminated to participants through  
19  
20 public education during follow-up.  
21  
22  
23  
24  
25

### 26 27 **Clinical outcomes**

28  
29 The primary study end point is postoperative recurrence following AF surgery defined  
30  
31 as persistence or recurrence of AF symptoms, or the development of recurrent  
32  
33 perianal sepsis or chronic AF within 3 to 6 months of surgery.<sup>23 18</sup> The second end  
34  
35 point is a composite outcome of postoperative comorbidities or any of the following  
36  
37 equivalent events including AFR, wound hemorrhage, fecal impaction, urinary  
38  
39 retention, delayed wound healing or unplanned hospitalization associated with AF  
40  
41 surgery. Outcomes were ascertained by the treating clinician combined with  
42  
43 outpatient medical records or patient self-reports.  
44  
45  
46  
47  
48  
49  
50

### 51 52 53 **Selection of predictor variables**

54  
55 Candidate variables for the prediction model of the composite outcome of  
56  
57 postoperative comorbidities will be screened according to the following pre-set  
58  
59  
60

1  
2  
3  
4 criteria: (1) prior clinical knowledge; (2) results from a systematic review updated in  
5  
6 November 2019 based on our published one<sup>4</sup> with sufficient evidence to include them  
7  
8 as predictive variables in the risk model for AFR as is demonstrated below; or (3)  
9  
10 agreed upon by a group of anorectal surgeons or experts for their clinical relevance  
11  
12 using a two-round Delphi survey. We initially identified the following covariates as  
13  
14 relevant candidate variables based on the systematic reviews as well as clinical  
15  
16 knowledge and/or relevance. The determination of all other candidate variables are  
17  
18 based on results of post-hoc analysis using univariable or multivariable survival  
19  
20 analyses with a threshold of  $p < 0.05$ .  
21  
22  
23  
24  
25  
26  
27  
28  
29

30 Factors identified from the systematic reviews and Delphi survey (manuscript under  
31  
32 review), will be measured at baseline. These include factors involving the identified  
33  
34 significant risk factors which are reported in our meta-analysis are presented as  
35  
36 follows:  
37  
38

- 39 ▶ Prior anal surgery.
- 40
- 41 ▶ Seton placement surgery.
- 42
- 43 ▶ High transsphincteric fistula.
- 44
- 45 ▶ Internal opening unidentified.
- 46
- 47 ▶ Horseshoe extensions.
- 48
- 49 ▶ Multiple fistula tracts.
- 50
- 51
- 52
- 53
- 54
- 55
- 56
- 57
- 58
- 59
- 60



1  
2  
3  
4 Some of the demographic factors and surgery details will also be collected due to  
5  
6 limited power in the literature reviews and some non-significant potential factors (eg,  
7  
8 smoking or alcohol use) may be risk factors and are also included as follows:  
9  
10

- 11 ▶ Gender.
- 12
- 13 ▶ Age.
- 14
- 15
- 16 ▶ Smoking.
- 17
- 18
- 19 ▶ Alcohol use.
- 20
- 21
- 22 ▶ Diabetes mellitus.
- 23
- 24
- 25 ▶ Obesity.
- 26
- 27 ▶ Preoperative seton drainage.
- 28
- 29
- 30 ▶ High internal opening.
- 31
- 32 ▶ Postoperative drainage.
- 33
- 34
- 35 ▶ Suprlevator extensions.
- 36

37  
38 Other factors like laboratory examinations and MR imaging parameters (height of the  
39  
40 internal openings, height and number of fistula, etc.) will also be collected. Moreover,  
41  
42 some other factors like chronic steroid therapy, diverting stoma, the surgeon's level of  
43  
44 training, postoperative bowel confinement and antibiotic prophylaxis reported in  
45  
46 previous literature are selected for regression analysis as well. In addition, relevant  
47  
48 factors from the expert-opinion survey were also assessed including the number of  
49  
50 prior anal fistula surgeries, types of surgery performed such as staged fistulotomies,  
51  
52  
53  
54  
55  
56 endorectal advancement flap and ligation of the intersphincteric fistula tract with and  
57  
58  
59  
60

1  
2  
3  
4 without seton drains, some nutrition parameters and immunomodulation medication  
5  
6 use.  
7  
8  
9

### 10 11 **Categorization of potential predictors** 12

13  
14 For categorical predictors, we can code them as “factor” variables, with coding as  
15  
16 dummy variables, for example, smoking is coded originally as “1” for never smoker,  
17  
18 “2” for past smoker, and “3” for current smoker and never smoker was selected as the  
19  
20 reference category. Similar manner can be applied with alcohol use.<sup>19</sup> Continuous  
21  
22 variables formally should be measured at an interval or ratio scale, and should be able  
23  
24 to take any value in a range. We treat ordered variables as linear which is generally  
25  
26 reasonable for prediction. In other cases, continuous predictors can be grouped with  
27  
28 meaningful categorization; for example, body mass index (BMI) can be classified  
29  
30 based on internationally recognized categories (i.e., underweight, normal weight,  
31  
32 overweight, and obesity).<sup>20</sup> Based on previous experiences, we will be deriving some  
33  
34 predictors based on the responses of the surveys. However, in case some subjectivity  
35  
36 in the classifications of these predictors may occur, sensitivity analyses will be  
37  
38 performed to examine the robustness of our definitions during model development  
39  
40 and validation.  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52

### 53 **Study quality control for the prediction models** 54

55  
56 Based on the summary of methodological quality and developmental stage of  
57  
58 prediction models by van Oort et al., we are developing predesigned criteria for  
59  
60

1  
2  
3  
4 quality control of our prediction models, which can make us carry out the study and  
5  
6 report the results more rigorously.<sup>21 22</sup> The methodological checklist of the study are  
7  
8 presented in supplementary material.  
9  
10

### 11 12 13 14 **Missing data**

15  
16 Candidate predictors with more than 60% missingness will be excluded. For those  
17  
18 less than 60% missingness, multiple imputation are to be performed by imputing 20  
19  
20 complete data sets using multivariate normal regression,<sup>23-26</sup> which can reasonably  
21  
22 approximate the true distributional relationship between the missing values and the  
23  
24 available ones.<sup>27</sup> Among various multiple imputation approaches, fully conditional  
25  
26 specification (FCS) and multivariate normal imputation (MVNI) are preferred,  
27  
28 because they have been proved to be generally less biased than complete-case analysis.  
29  
30 They can both generate similar results in the presence of either binary or ordinal  
31  
32 variables that are not generally normally distributed.<sup>27</sup>  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42

### 43 **Statistical analysis for model derivation**

44  
45 Logistic regression will be applied to develop our prediction models for the binary  
46  
47 outcomes. All data processing and statistical analysis will be performed using  
48  
49 EmpowerStats software ([www. empowerstats.com](http://www.empowerstats.com); X&Y Solutions, Inc., Boston, MA,  
50  
51 USA) and statistical software packages R (R Foundation, Vienna, Austria).  
52  
53 We will first study the association between each potential variable and the outcome  
54  
55 based on univariable analysis. Variables are considered further for multivariable  
56  
57  
58  
59  
60

1  
2  
3  
4 regression modelling when they are associated with a p-value less than 0.20.  
5

6 Normality or linearity will be evaluated for the continuous predictors. Fractional  
7  
8  
9 polynomials are advocated for associations between the continuous predictors and the  
10  
11 outcome for non-linear relationships.<sup>28 29</sup> We will perform backward stepwise  
12  
13 selection with a  $p < 0.001$  as the inclusion threshold and a  $p > 0.05$  as the exclusion  
14  
15 threshold for each imputed data set.  
16  
17  
18

19  
20  
21  
22 Predictors which appear in the imputation models with an inclusion fraction of  $\geq 50\%$   
23  
24 are qualified for the final multivariable model. Though there is no consensus  
25  
26 regarding the optimal method for selecting predictors for inclusion, backwards  
27  
28 elimination is generally considered as the preferred procedure as reported by Mantel  
29  
30 et al.<sup>30</sup> Forward stepwise procedure will also be performed to repeat the analysis to  
31  
32 test the robustness of the models. Overall regression coefficient estimates of the  
33  
34 models will be generated with the combination of the imputed datasets based on  
35  
36 Rubin's Rules, while taking into account uncertainty in the imputed values.<sup>23 24 26</sup>  
37  
38 Collinearity will also be assessed which refers to the fact that predictors can have  
39  
40 strong correlation with each other, defined as correlation coefficient  $>0.8$ , or variance  
41  
42 inflation factor  $>10$ .<sup>31</sup> Then we will examine the interactions among the regression  
43  
44 models.  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55

## 56 **Prediction model performance assessment**

57  
58  
59  
60

1  
2  
3  
4 Prediction models will be developed with a random sample of 60% of the AF cohort  
5  
6 as the derivation cohort, and then validated with the remaining sample of 40% of the  
7  
8 cohort as the validation cohort. The predictive performance in the derivation and  
9  
10 validation cohort will be evaluated and reported by examining measures of predictive  
11  
12 accuracy, discrimination and calibration. Nagelkerke's  $R^2$  and Brier score will be used  
13  
14 for the measurement of predictive accuracy.<sup>32 33</sup> The discriminative ability of the  
15  
16 prediction models are evaluated using several statistics, which are according to the  
17  
18 discriminative and calibration ability in both derivation and validation AF cohort.  
19  
20 Model discrimination means the ability of the models to differentiate between  
21  
22 high-risk patients and low-risk patients (having high or low risk of AFR or  
23  
24 surgery-related complications). This will be assessed via Harrell's concordance  
25  
26 statistic (C-index).<sup>34</sup> The calculation of the C-index will be performed in each of the  
27  
28 20 imputed data sets, and then averaged based on Rubin's rule.<sup>35</sup> The model is  
29  
30 interpreted as having no discriminatory ability when a value of C-index is 0.5, and has  
31  
32 perfect discrimination when a value of C-index is 1.0.<sup>34</sup> Calibration implies the  
33  
34 agreement between the predicted outcomes and the observed outcomes, which is  
35  
36 evaluated with the Hosmer and Lemeshow test for goodness of fit in all imputed  
37  
38 datasets presented with calibration plots.<sup>36</sup> Calibration-in-the-large, which defines as  
39  
40 the agreement between mean observed outcomes and mean predictions, will also be  
41  
42 assessed for calibration.<sup>37</sup>  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

### **Internal and external validation of prediction model**

1  
2  
3  
4 To make the prediction models reproducible, we have to conduct internal validation.  
5  
6 Bootstrapping technique, as one of the most attractive resampling techniques, is a  
7  
8 mostly applied validation method, which seems to be most efficient for obtaining  
9  
10 stable optimism-corrected estimates.<sup>34 38</sup> It has been reported that bootstrap validation  
11  
12 is a feasible technique for most prediction models with at least a 500 bootstrap  
13  
14 resampling procedure using Harrell's validate function, which can adjust the  
15  
16 developed models for over-fitting.<sup>39</sup> We will also apply temporal validation as  
17  
18 external validation using a more recent AF patient cohort.<sup>40</sup>  
19  
20  
21  
22  
23  
24  
25  
26

### 27 **Sample size**

28  
29 Since there are no widely accepted methods for the estimation of the sample size  
30  
31 requirements to develop the risk prediction models, the size of this AF cohort will be  
32  
33 calculated to have 20 events per candidate predictive variable (EPV, defined as the  
34  
35 ratio of the number of individuals with the outcome event to the number of candidate  
36  
37 predictors), which can generally eliminate bias in regression coefficients for  
38  
39 prediction models with low-prevalence binary predictor development (the estimated  
40  
41 recurrence rate <20%) and adequately power the logistic regression models.<sup>41,42</sup>  
42  
43  
44  
45  
46  
47

48 According the findings by Ogundimu et al.<sup>43</sup>, a higher EPV ( $\geq 20$ ) can generally  
49  
50 eliminate bias in regression coefficients for prediction models with low-prevalence  
51  
52 binary predictors development. Then we estimate 400 events allow for 20 predictor  
53  
54 variables (EPV=20). Considering 5-20% recurrence rate, we assume that at least  
55  
56 4000-8000 patients should be collected for model development. In addition, surgery  
57  
58  
59  
60

1  
2  
3  
4 type and fistula type-stratified analyses will also be performed to examine the  
5  
6 different effect of these factors on disease recurrence or other related complications in  
7  
8 each subgroup. Risk prediction models, if possible, can also be developed in those  
9  
10 sub-populations. The cohort size with more than 4000-8000 patients will provide  
11  
12 sufficient power to perform those analyses and develop prediction models in those  
13  
14  
15  
16  
17 subgroups.

### 22 **Follow-up and methodological quality control**

23  
24 The application of WeChat questionnaires to collect data will inevitably increase the  
25  
26 probability of missing data. However, we have made some pre-designed  
27  
28 countermeasures. For example, we have set up follow-up reminders via the WeChat  
29  
30 questionnaire system. Moreover, every week two trained clinical fellows cross-check  
31  
32 the data, and will contact the respondents by phone or WeChat about the missing  
33  
34 contents, which can minimize the missing data and lost to follow-up rate.  
35  
36  
37  
38  
39

40  
41  
42 Limited studies have identified specific criteria for quality control in a prediction  
43  
44 model, but we have strictly adhered to the guidelines for the reporting of studies  
45  
46 developing, validating multivariable clinical prediction models as is reported in the  
47  
48 TRIPOD (Transparent Reporting of a multivariable prediction model for Individual  
49  
50 Prognosis Or Diagnosis) Statement to ensure methodological rigour.<sup>44</sup> All issues have  
51  
52 been addressed in this study in supplementary material.  
53  
54  
55  
56  
57  
58  
59  
60

### **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 recurrence 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,<sup>45-47</sup> lifestyle factors such as smoking, alcohol abuse,<sup>48 49</sup> 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<sup>48 50 51</sup>) easily ascertainable before surgery may predict AFR. Similar risk prediction models exist in other



1  
2  
3  
4 diseases, such as the Framingham Risk Score model to predict cardiovascular disease  
5  
6 risk<sup>52</sup> and the Korean Crohn's Disease Prediction (KCDP) model to predict the  
7  
8 clinical course of Crohn's disease.<sup>53</sup> Until now, risk factor investigation of predictors  
9  
10 of perioperative surgery-related complications has been limited to assessment of  
11  
12 single predictors with small sample size. The risk prediction models can help inform  
13  
14 surgeons regarding high risk AF patients based on the overall risk factors. The  
15  
16 primary purpose of study was to develop two risk prediction models to facilitate  
17  
18 surgeons in identifying AF patients who will have surgical treatment at higher risk of  
19  
20 developing recurrence and surgery-related complications. The predictive models will  
21  
22 help both clinicians and patients identify the risk of complications after AF surgery in  
23  
24 advance, take necessary interventions to reduce the risk of surgery-related  
25  
26 complications as well as the personal and social financial burden brought about by  
27  
28 those complications. The accurate risk prediction models are especially instructive for  
29  
30 the development of the optimal surgical plan to achieve optimal surgical outcome.  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42

43 Our study has several strengths. To the best of our knowledge, this is the first study  
44  
45 with the primary aim to develop, internally and externally validate multivariable  
46  
47 prediction models for AFR and related complications following AF surgery.  
48  
49

50 Multidimensional clinically useful candidate predictors will be fully examined from a  
51  
52 variety of sources including our published and updated systematic reviews, expert  
53  
54 opinions from Delphi surveys and univariable or multivariable logistic regression  
55  
56 analysis. Second, we will apply the internal and external validation of the prediction  
57  
58  
59  
60

1  
2  
3  
4 models using bootstrapping procedure. Third, multiple imputation will also be used to  
5  
6 treat the missing data. Last but not least, our study is a prospective cohort one with  
7  
8 adequate follow-up period which can minimize certain bias.  
9  
10

11  
12  
13  
14 Our study also has limitations. As many of the variables are collected through a  
15  
16 WeChat questionnaire system during the hospitalization and follow-up, a higher  
17  
18 probability of missing data due to non-response bias may occur. To solve this issue,  
19  
20 we regularly send reminders to those who does not respond after discharge. Second,  
21  
22 though we will investigate a series of potential predictors, some more potentially  
23  
24 predictors will not be involved or not collected in the current study, data related to  
25  
26 postoperative nursing strategy and outpatient follow-up frequency for example.  
27  
28 Moreover, bias may also result from the single-center recruitment of our study and  
29  
30 will be improved through multicenter recruitment in the future. Last but not least, one  
31  
32 key potential issue that needs to be considered a priori is the variable recurrence rate  
33  
34 depending on the risk factors included and identified in the testing and validating  
35  
36 cohorts since this may affect the C-index. It is possible that a lack of key variable  
37  
38 inclusions in the models may result in decreased discriminatory ability. This is a  
39  
40 function of the database and points of interest included that may need to be  
41  
42 maximized before proceeding with development of the testing model.  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55

56 The newly developed risk algorithms may have significant applications in clinical  
57  
58 practice by helping recommend optimal surgical approach for a specific AF patient,  
59  
60

1  
2  
3  
4 as well as intensive perioperative care and education, timely assessment and  
5  
6 discussion of the need for interventions to those most at high risk of developing  
7  
8 recurrence or surgery-related complications. The models will specifically identify AF  
9  
10 patients who are likely to develop recurrence or related complications following AF  
11  
12 surgery to offer the quantitative evaluation of the risk. Moreover, the models will also  
13  
14 provide reference information for preventing recurrence and reducing the rate of  
15  
16 recurrence after operation, and to intervene some high risk factors in the early stage.  
17  
18  
19  
20  
21  
22  
23  
24

25 In summary, this study protocol summarizes the design of development and validation  
26  
27 studies for a risk screening tool in patients receiving AF surgery. Results from this  
28  
29 study will be interpreted for the purpose of clinical decision making. The models to be  
30  
31 developed of the study could be used to make new recommendations for perioperative  
32  
33 AF patients.  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

## References

1. Tantiplachiva 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.ijvsu.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

- 1  
2  
3  
4 7. Gottgens KW, Janssen PT, Heemskerk J, et al. Long-term outcome of low perianal  
5  
6 fistulas treated by fistulotomy: a multicenter study. *International journal of*  
7  
8 *colorectal disease* 2015;30(2):213-9. doi: 10.1007/s00384-014-2072-y  
9  
10  
11  
12 8. Bakhtawar N, Usman M. Factors Increasing the Risk of Recurrence in Fistula-in-ano.  
13  
14 *Cureus* 2019;11(3):e4200. doi: 10.7759/cureus.4200  
15  
16  
17  
18 9. Sygut A, Mik M, Trzcinski R, et al. How the location of the internal opening of anal  
19  
20 fistulas affect the treatment results of primary transsphincteric fistulas.  
21  
22 *Langenbeck's archives of surgery* 2010;395(8):1055-9. doi:  
23  
24 10.1007/s00423-009-0562-0  
25  
26  
27  
28  
29  
30 10. Dziki A, Bartos M. Seton treatment of anal fistula: experience with a new  
31  
32 modification. *The European journal of surgery = Acta chirurgica* 1998;164(7):543-8.  
33  
34 doi: 10.1080/110241598750005930  
35  
36  
37  
38  
39 11. Hermann J, Eder P, Banasiewicz T, et al. Current management of anal fistulas in  
40  
41 Crohn's disease. *Przegląd gastroenterologiczny* 2015;10(2):83-8. doi:  
42  
43 10.5114/pg.2015.49684  
44  
45  
46  
47  
48 12. Hamadani A, Haigh PI, Liu IL, et al. Who is at risk for developing chronic anal  
49  
50 fistula or recurrent anal sepsis after initial perianal abscess? *Diseases of the colon*  
51  
52 *and rectum* 2009;52(2):217-21. doi: 10.1007/DCR.0b013e31819a5c52  
53  
54  
55  
56  
57  
58  
59  
60

- 1  
2  
3  
4 13. Ellis CN, Clark S. Effect of tobacco smoking on advancement flap repair of  
5  
6 complex anal fistulas. *Diseases of the colon and rectum* 2007;50(4):459-63. doi:  
7  
8 10.1007/s10350-006-0829-2  
9  
10  
11  
12 14. Moons KG, Altman DG, Vergouwe Y, et al. Prognosis and prognostic research:  
13  
14 application and impact of prognostic models in clinical practice. *Bmj* 2009;338:b606.  
15  
16 doi: 10.1136/bmj.b606  
17  
18  
19  
20 15. Ming C, Viassolo V, Probst-Hensch N, et al. Machine learning techniques for  
21  
22 personalized breast cancer risk prediction: comparison with the BCRAT and  
23  
24 BOADICEA models. *Breast cancer research : BCR* 2019;21(1):75. doi:  
25  
26 10.1186/s13058-019-1158-4  
27  
28  
29  
30  
31  
32 16. Khan SS, Ning H, Shah SJ, et al. 10-Year Risk Equations for Incident Heart Failure  
33  
34 in the General Population. *Journal of the American College of Cardiology*  
35  
36 2019;73(19):2388-97. doi: 10.1016/j.jacc.2019.02.057  
37  
38  
39  
40  
41 17. Terry MB, Liao Y, Whittemore AS, et al. 10-year performance of four models of  
42  
43 breast cancer risk: a validation study. *The Lancet Oncology* 2019;20(4):504-17. doi:  
44  
45 10.1016/S1470-2045(18)30902-1  
46  
47  
48  
49 18. Yano T, Asano M, Matsuda Y, et al. Prognostic factors for recurrence following  
50  
51 the initial drainage of an anorectal abscess. *International journal of colorectal*  
52  
53 *disease* 2010;25(12):1495-8. doi: 10.1007/s00384-010-1011-9 [published Online  
54  
55 First: 2010/07/20]  
56  
57  
58  
59  
60

1  
2  
3  
4 19. Steyerberg EW. Coding of Categorical and Continuous Predictors. In: Steyerberg  
5  
6 EW, ed. *Clinical Prediction Models: A Practical Approach to Development, Validation,*  
7  
8 *and Updating.* Cham: Springer International Publishing 2019:175-90.

9  
10  
11  
12 20. Morgan TM, Elashoff RM. Effect of Categorizing a Continuous Covariate on the  
13  
14  
15 Comparison of Survival Time. *Journal of the American Statistical Association*  
16  
17  
18 1986;81(396):917-21. doi: 10.1080/01621459.1986.10478350

19  
20  
21 21. van Oort L, van den Berg T, Koes BW, et al. Preliminary state of development of  
22  
23  
24 prediction models for primary care physical therapy: a systematic review. *Journal of*  
25  
26  
27 *Clinical Epidemiology* 2012;65(12):1257-66. doi:  
28  
29 <https://doi.org/10.1016/j.jclinepi.2012.05.007>

30  
31  
32  
33 22. Rushton AB, Verra ML, Emms A, et al. Development and validation of two clinical  
34  
35  
36 prediction models to inform clinical decision-making for lumbar spinal fusion surgery  
37  
38  
39 for degenerative disorders and rehabilitation following surgery: protocol for a  
40  
41  
42 prospective observational study. *BMJ Open* 2018;8(5):e021078. doi:  
43  
44  
45 [10.1136/bmjopen-2017-021078](https://doi.org/10.1136/bmjopen-2017-021078)

46  
47 23. Wood AM, White IR, Royston P. How should variable selection be performed  
48  
49  
50 with multiply imputed data? *Statistics in Medicine* 2008;27(17):3227-46. doi:  
51  
52  
53 [10.1002/sim.3177](https://doi.org/10.1002/sim.3177)

- 1  
2  
3  
4 24. White IR, Royston P, Wood AM. Multiple imputation using chained equations:  
5  
6 Issues and guidance for practice. *Statistics in Medicine* 2011;30(4):377-99. doi:  
7  
8 10.1002/sim.4067  
9  
10  
11  
12 25. Schafer JL. Multiple imputation: a primer. *Statistical Methods in Medical*  
13  
14 *Research* 1999;8(1):3-15. doi: 10.1177/096228029900800102  
15  
16  
17  
18 26. Marshall A, Altman DG, Holder RL, et al. Combining estimates of interest in  
19  
20 prognostic modelling studies after multiple imputation: current practice and  
21  
22 guidelines. *BMC Medical Research Methodology* 2009;9(1):57. doi:  
23  
24 10.1186/1471-2288-9-57  
25  
26  
27  
28  
29 27. Lee KJ, Carlin JB. Multiple Imputation for Missing Data: Fully Conditional  
30  
31 Specification Versus Multivariate Normal Imputation. *American Journal of*  
32  
33 *Epidemiology* 2010;171(5):624-32. doi: 10.1093/aje/kwp425  
34  
35  
36  
37  
38 28. Sauerbrei W, Royston P, Binder H. Selection of important variables and  
39  
40 determination of functional form for continuous predictors in multivariable model  
41  
42 building. *Statistics in Medicine* 2007;26(30):5512-28. doi: 10.1002/sim.3148  
43  
44  
45  
46  
47 29. Royston P, Sauerbrei W. Building multivariable regression models with  
48  
49 continuous covariates in clinical epidemiology--with an emphasis on fractional  
50  
51 polynomials. *Methods of information in medicine* 2005;44(4):561-71.  
52  
53  
54  
55  
56 30. Mantel N. Why Stepdown Procedures in Variable Selection. *Technometrics*  
57  
58 1970;12(3):621-25. doi: 10.1080/00401706.1970.10488701  
59  
60



- 1  
2  
3  
4 31. Vittinghoff E, Glidden DV, Shiboski SC, et al. Regression methods in biostatistics:  
5  
6 linear, logistic, survival, and repeated measures models: Springer Science & Business  
7  
8 Media 2011.  
9  
10  
11  
12 32. Steyerberg EW, Vickers AJ, Cook NR, et al. Assessing the performance of  
13  
14 prediction models: a framework for traditional and novel measures. *Epidemiology*  
15  
16 2010;21(1):128-38. doi: 10.1097/EDE.0b013e3181c30fb2  
17  
18  
19  
20  
21 33. NAGELKERKE NJD. A note on a general definition of the coefficient of  
22  
23 determination. *Biometrika* 1991;78(3):691-92. doi: 10.1093/biomet/78.3.691  
24  
25  
26  
27 34. Harrell Jr FE. Regression modeling strategies: with applications to linear models,  
28  
29 logistic and ordinal regression, and survival analysis: Springer 2015.  
30  
31  
32  
33 35. Rubin DB. Multiple imputation for nonresponse in surveys: John Wiley & Sons  
34  
35  
36 2004.  
37  
38  
39  
40 36. Hosmer Jr DW, Lemeshow S, Sturdivant RX. Applied logistic regression: John  
41  
42 Wiley & Sons 2013.  
43  
44  
45  
46 37. Snell KI, Ensor J, Debray TP, et al. Meta-analysis of prediction model performance  
47  
48 across multiple studies: Which scale helps ensure between-study normality for the  
49  
50 C-statistic and calibration measures? *Statistical methods in medical research*  
51  
52 2018;27(11):3505-22.  
53  
54  
55  
56  
57  
58  
59  
60

- 1  
2  
3  
4 38. Chatfield C. Model uncertainty, data mining and statistical inference. *Journal of*  
5  
6 the Royal Statistical Society: Series A (Statistics in Society) 1995;158(3):419-44.  
7  
8  
9  
10 39. Steyerberg EW, Bleeker SE, Moll HA, et al. Internal and external validation of  
11  
12 predictive models: a simulation study of bias and precision in small samples. *Journal*  
13  
14 of clinical epidemiology 2003;56(5):441-47.  
15  
16  
17  
18 40. Justice AC, Covinsky KE, Berlin JA. Assessing the generalizability of prognostic  
19  
20 information. *Annals of internal medicine* 1999;130(6):515-24.  
21  
22  
23  
24 41. Concato J, Peduzzi P, Holford TR, et al. Importance of events per independent  
25  
26 variable in proportional hazards analysis I. Background, goals, and general strategy.  
27  
28 *Journal of clinical epidemiology* 1995;48(12):1495-501.  
29  
30  
31  
32  
33 42. Steyerberg EW. *Clinical Prediction Models: A Practical Approach to Development,*  
34  
35 *Validation, and Updating.* SPRINGER NATURE, 2019.  
36  
37  
38  
39 43. Ogundimu EO, Altman DG, Collins GS. Adequate sample size for developing  
40  
41 prediction models is not simply related to events per variable. *J Clin Epidemiol.*  
42  
43 2016;76:175–182.  
44  
45  
46  
47  
48 44. Moons KG, Altman DG, Reitsma JB, et al. Transparent Reporting of a  
49  
50 multivariable prediction model for Individual Prognosis or Diagnosis (TRIPOD):  
51  
52 explanation and elaboration. *Annals of internal medicine* 2015;162(1):W1-W73.  
53  
54  
55  
56  
57  
58  
59  
60

- 1  
2  
3  
4 45. Hamadani A, Haigh PI, Liu I-LA, et al. Who is at risk for developing chronic anal  
5  
6 fistula or recurrent anal sepsis after initial perianal abscess? *Diseases of the colon &*  
7  
8 *rectum* 2009;52(2):217-21.  
9  
10  
11  
12 46. Sonoda T, Hull T, Piedmonte MR, et al. Outcomes of primary repair of anorectal  
13  
14 and rectovaginal fistulas using the endorectal advancement flap. *Diseases of the*  
15  
16 *colon & rectum* 2002;45(12):1622-28.  
17  
18  
19  
20 47. Hyman N, O'Brien S, Osler T. Outcomes after fistulotomy: results of a prospective,  
21  
22 multicenter regional study. *Diseases of the Colon & Rectum* 2009;52(12):2022-27.  
23  
24  
25  
26 48. Ellis CN, Clark S. Effect of tobacco smoking on advancement flap repair of  
27  
28 complex anal fistulas. *Diseases of the colon & rectum* 2007;50(4):459-63.  
29  
30  
31  
32 49. Schwandner T, Roblick MH, Kierer W, et al. Surgical treatment of complex anal  
33  
34 fistulas with the anal fistula plug: a prospective, multicenter study. *Diseases of the*  
35  
36 *Colon & Rectum* 2009;52(9):1578-83.  
37  
38  
39  
40 50. Garcia-Aguilar J, Belmonte C, Wong WD, et al. Anal fistula surgery: Factors  
41  
42 associated with recurrence and incontinence. *Diseases of the colon & rectum*  
43  
44 1996;39(7):723-29.  
45  
46  
47  
48 51. Athanasiadis S, Köhler A, Nafe M. Treatment of high anal fistulae by primary  
49  
50 occlusion of the internal ostium, drainage of the intersphincteric space, and mucosal  
51  
52 advancement flap. *International journal of colorectal disease* 1994;9(3):153-57.  
53  
54  
55  
56  
57  
58  
59  
60

- 1  
2  
3  
4 52. Fox ER, Samdarshi TE, Musani SK, et al. Development and validation of risk  
5  
6 prediction models for cardiovascular events in black adults: the Jackson Heart Study  
7  
8 Cohort. JAMA cardiology 2016;1(1):15-25.  
9  
10  
11  
12 53. Park Y, Cheon JH, Park YL, et al. Development of a novel predictive model for the  
13  
14 clinical course of Crohn's disease: results from the connect study. Inflammatory  
15  
16 bowel diseases 2017;23(7):1071-79.  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

1  
2  
3 **Figure legend**  
4  
5

6 Figure 1. Flowchart of prediction model development and assessment.  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

For peer review only

1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

Figure 1: Flowchart of prediction model development and assessment.

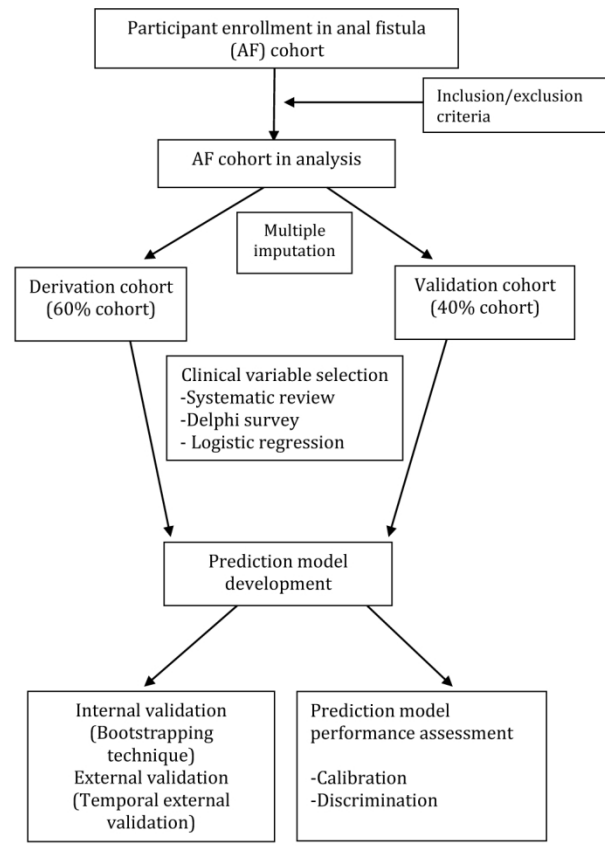


Figure 1. Flowchart of prediction model development and assessment.



## 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
<b>Title and abstract</b>				
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
<b>Introduction</b>				
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
<b>Methods</b>				
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
<b>Results</b>				
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	-

## TRIPOD Checklist: Prediction Model Development and Validation

specification	15b	D	coefficients, and model intercept or baseline survival at a given time point). 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).	-
<b>Discussion</b>				
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
<b>Other information</b>				
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.





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
<b>Title and abstract</b>				
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
<b>Introduction</b>				
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
<b>Methods</b>				
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
<b>Results</b>				
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	-

## TRIPOD Checklist: Prediction Model Development and Validation

specification	15b	D	coefficients, and model intercept or baseline survival at a given time point). Explain how to 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).	-
<b>Discussion</b>				
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
<b>Other information</b>				
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.

# 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:	<i>BMJ Open</i>
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
<b>Primary Subject Heading</b>:	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



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.

1  
2  
3  
4 **Development of screening tools to predict the risk of recurrence and related**  
5 **complications following anal fistula surgery: protocol for a prospective cohort**  
6 **study**  
7  
8  
9

10  
11 Zubing Mei<sup>1,2\*</sup>; Yue Li<sup>1\*</sup>; Zhijun Zhang<sup>1</sup>; Haikun Zhou<sup>1</sup>; Suzhi Liu<sup>1</sup>; Ye Han<sup>1</sup>;  
12 Peixin Du<sup>1</sup>; Xiufang Qin<sup>3</sup>; Zhuo Shao<sup>4</sup>; Maojun Ge<sup>5</sup>; Qingming Wang<sup>1</sup>; Wei Yang<sup>1,2</sup>

13  
14  
15 \*Contributed equally to this manuscript  
16

- 17  
18  
19 1. Department of Anorectal Surgery, Shuguang Hospital, Shanghai University of  
20 Traditional Chinese Medicine, Shanghai, China;  
21  
22 2. Anorectal Disease Institute of Shuguang Hospital, Shanghai, China;  
23  
24 3. Department of Nursing, Shuguang Hospital, Shanghai University of  
25 Traditional Chinese Medicine, Shanghai, China;  
26  
27 4. Department of General Surgery, Changhai Hospital, The Second Military  
28 Medical University, Shanghai, China;  
29  
30 5. Department of General Surgery, Shuguang Hospital, Shanghai University of  
31 Traditional Chinese Medicine, Shanghai, China.  
32  
33  
34  
35

36  
37  
38 **Correspondence to:**

39 Dr. Zubing Mei, MD, PhD

40  
41 Department of Anorectal Surgery, Shuguang Hospital, Shanghai University of  
42 Traditional Chinese Medicine; Anorectal Disease Institute of Shuguang Hospital  
43 528 Zhangheng Road, Shanghai 201203, China  
44  
45

46 Tel: 086-2120256187

47  
48 E-mail: herrmayor@126.com  
49  
50

51  
52  
53 Professor. Wei Yang, MD

54 Department of Anorectal Surgery, Shuguang Hospital, Shanghai University of  
55 Traditional Chinese Medicine; Anorectal Disease Institute of Shuguang Hospital  
56 528 Zhangheng Road, Shanghai 201203, China  
57  
58  
59  
60

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. zyz-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. shslczdk04302), Shanghai Famous Old Traditional Chinese Medicine Experts Inheritance Studio Construction Program (grant no. SHGZS-2017028),

1  
2  
3  
4 Famous Old Traditional Chinese Medicine Experts Inheritance Studio Construction  
5  
6 Program of Shanghai University of TCM (grant no.SZYMZYGZS4016-2017028).  
7  
8  
9  
10

11  
12 **Abbreviations:** AF, anal fistula; AFR, anal fistula recurrence; BMI, body mass index;  
13  
14 FCS, fully conditional specification; MVNI, multivariate normal imputation; TRIPOD,  
15  
16 Transparent Reporting of a multivariable prediction model for Individual Prognosis  
17  
18 Or Diagnosis.  
19  
20  
21  
22

23  
24  
25 **Author contributions:**

26  
27  
28 Dr. Zubing Mei had full access to all of the data in the study and takes responsibility  
29  
30 for the integrity of the data and the accuracy of the data analysis. Drs Mei and Li are  
31  
32 co-first authors of this article.  
33

34  
35  
36 Study concept and design: Zubing Mei.

37  
38  
39 Acquisition, analysis, or interpretation of data: Yue Li, Zubing Mei, Zhijun Zhang, Ye  
40  
41 Han, Suzhi Liu, Haikun Zhou, Peixin Du, Xiufang Qin, Qingming Wang, Wei Yang.  
42

43  
44 Drafting of the manuscript: Zubing Mei.

45  
46  
47 Critical revision of the manuscript for important intellectual content: All authors.

48  
49  
50 Statistical analysis: Zubing Mei, Zhuo Shao, Maojun Ge.

51  
52  
53 Administrative, technical, or material support: All authors.

54  
55  
56 Study supervision: Zubing Mei.  
57  
58

59  
60 **Competing interests statement:**

1  
2  
3  
4 The authors declare that they have no conflict of interest.  
5  
6  
7  
8

9 **Role of the Funder/Sponsor:**  
10

11 The funder of the study had no role in the study design, data collection, data analysis,  
12  
13 data interpretation or writing of the manuscript. The corresponding author had full  
14  
15 access to all the data in the study and has final responsibility for the decision to  
16  
17 submit for publication.  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60



## **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

1  
2  
3  
4 warn, motivate and empower patients to avoid some modifiable risk factors to prevent  
5  
6 postoperative complications early. This study will also provide alternative tools for  
7  
8 the early screening of high-risk patients with AFR and related complications, helping  
9  
10 surgeons better understand the aetiology and outcomes of AF in an earlier stage.  
11  
12  
13  
14  
15  
16

### 17 **Ethics and dissemination**

18  
19 The study was approved by the Institutional Review Board of Shuguang Hospital  
20  
21 affiliated with Shanghai University of Traditional Chinese Medicine (approval  
22  
23 number: 2019-699-54-01). The results of this study will be submitted to international  
24  
25 scientific peer-reviewed journals or conferences in surgery, anorectal surgery or  
26  
27 anorectal diseases.  
28  
29  
30  
31  
32  
33  
34

### 35 **Trial registration number**

36  
37 Chinese Clinical Trial Registry (ChiCTR1900025069); Pre-results.  
38  
39  
40  
41  
42

### 43 **Key words**

44  
45 Anal fistula; treatment outcome; recurrence; surgery; cohort study; prediction model  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

### 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 ( $\geq 20$ ) 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.

## **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.<sup>1</sup> 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,<sup>2,3</sup> 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.<sup>4</sup> 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%.<sup>5-8</sup> 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.<sup>9,10</sup> 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.<sup>4,11-13</sup> Li et.al retrospectively analysed 1783 patients with AF receiving surgical treatment and found that the location of AF,

1  
2  
3  
4 previous perianal surgery, seton history and enteritis were independent risk factors for  
5  
6 AFR.<sup>5</sup> Recently, according to the evidence grading criteria based on Egger's P value,  
7  
8  
9 total sample size and between-study heterogeneity, we published a meta-analysis  
10  
11 involving 20 studies with 6168 patients and concluded that high transsphincteric  
12  
13 fistula, unidentified internal opening, and horseshoe extensions were independent risk  
14  
15 factors for AFR with high-quality evidence, while prior anal surgery, seton placement  
16  
17 surgery, and multiple fistula tracts were demonstrated to be risk factors for AFR with  
18  
19 moderate-quality evidence.<sup>4</sup>  
20  
21  
22  
23  
24  
25  
26

27 Factors influencing other perioperative complications related AF surgery  
28  
29 including wound haemorrhage, faecal impaction, urinary retention, delayed wound  
30  
31 healing and unplanned hospitalization are also rarely reported. Therefore, there is an  
32  
33 urgent need to develop risk prediction tools for the complete profile of risk factors for  
34  
35 AFR and related complications.  
36  
37  
38  
39  
40  
41  
42

43 Which AF patients will be cured after surgery and which ones will not, are  
44  
45 rarely investigated. The development of a prediction model for AFR following  
46  
47 surgery to identify those patients with a higher risk of developing complications  
48  
49 during follow up, would be of significant importance. First, surgeons can provide  
50  
51 patients preoperatively with an estimated surgical cure rate according to the prediction  
52  
53 models. Moreover, the current knowledge in the literature reporting potential  
54  
55 predictive factors could help patients become familiar their individual risk factors and  
56  
57  
58  
59  
60

1  
2  
3  
4 avoid modifiable ones to improve the cure rate, which has been well described and  
5  
6 applied in the prevention of other diseases.<sup>14-17</sup>  
7  
8  
9  
10

11  
12 However, to date, there are no effective screening tools to evaluate and predict  
13  
14 the risk of recurrence or other adverse outcomes of AF. Therefore, the aim of the  
15  
16 current study was to develop and validate multivariable prediction models that predict  
17  
18 postoperative AFR and related complications.  
19  
20  
21  
22

### 23 24 **Aims and objectives**

25  
26  
27 The aim of this study was to develop risk prediction models for postoperative  
28  
29 recurrence as well as other surgery-related complications in a prospective  
30  
31 hospital-based AF cohort. A risk prediction model for perioperative complications  
32  
33 will also be developed. A flowchart of prediction model development and assessment  
34  
35 is provided in Figure 1.  
36  
37  
38

39  
40 The detailed tasks of this study are as follows:

- 41  
42 1. Calculate the 3- to 6-month incidence of recurrence, and related  
43  
44 complications in patients following AF surgery.  
45  
46
- 47  
48 2. Establish the risk factors that significantly predict postoperative AFR and  
49  
50 related complications based on the AF cohort in a tertiary referral centre.  
51  
52
- 53  
54 3. Develop and validate the risk prediction models for postoperative AFR and  
55  
56 related complications.  
57  
58  
59  
60

1  
2  
3  
4 4. Considering the different scenarios for different surgical interventions,  
5  
6 conduct stratified analyses based on surgery type. If possible, a risk prediction model  
7  
8 will also be developed in relevant subpopulations, such as those only receiving  
9  
10 fistulectomy or fistulotomy, which can account for more than 60% of our AF cohort  
11  
12 populations.  
13  
14  
15

16  
17 We also examine the following two hypotheses:  
18

19  
20 1. Patient-related demographic characteristics, fistula and surgery-related  
21  
22 factors are predictive of postoperative AFR and related complications as dependent  
23  
24 variables.  
25

26  
27 2. The risk prediction models for postoperative AFR and related complications  
28  
29 developed in our study have more than 70% discriminating power.  
30  
31  
32  
33  
34

## 35 **Patients and Methods**

### 36 37 **Study design and participants**

38  
39 This study is a single-centre, prospective observational study on a hospital-based  
40  
41 cohort enrolled at a tertiary referral centre in Shanghai, China.  
42  
43  
44  
45  
46  
47

### 48 **Eligibility criteria**

49  
50 The enrolment of the cohort subjects was initiated in June, 2019. All subjects who  
51  
52 will undergo surgical intervention for AF will be included. All operations will be  
53  
54 performed by a group of colon and rectal surgeons at Shuguang Hospital, a regional  
55  
56 tertiary referral centre. The exclusion criteria were age < 18 years, noncryptoglandular  
57  
58  
59  
60

1  
2  
3  
4 fistula (e.g., anal fistula due to inflammatory bowel disease, human  
5  
6 immunodeficiency virus, malignant cancer, or obstetrical trauma), and rectovaginal or  
7  
8 rectourethral fistula. The electronic medical records of the included subjects were  
9  
10 complete.  
11  
12

13  
14  
15  
16  
17 Trained clinical investigators are collecting data in several categories, including  
18  
19 baseline demographics, laboratory examinations, clinical data, imaging findings and  
20  
21 follow-up information 3 to 6 months postoperatively. Planned clinical reviews or  
22  
23 electronic surveys are conducted during hospitalization and every 0.5 to 3 months  
24  
25 after discharge for 6 months.  
26  
27  
28  
29  
30  
31

### 32 **Data collection**

33  
34  
35 The research team comprised a principal investigator and 5 to 8 anorectal surgeons  
36  
37 who were trained and supervised by the Ethics Committee of Shuguang Hospital.  
38  
39  
40 Written informed consent was obtained from all patients. The investigators did not  
41  
42 intervene in any aspects of patient surveys at any stage of data collection and  
43  
44 follow-up. Data were collected using a convenient follow-up system supported by  
45  
46 Empower EDC (Solutions, Boston, Massachusetts, USA). This electronic system  
47  
48 introduces a machine learning algorithm, through which we can use the data already  
49  
50 entered in the Empower system to train the algorithm model and let the system itself  
51  
52 develop quality control algorithms, validate the entered data and identify missing or  
53  
54 suspicious data. Finally, the data manager will check the missing or suspicious data,  
55  
56  
57  
58  
59  
60



1  
2  
3  
4 confirm their completeness and ask the data manager to provide additional data when  
5  
6 necessary. Furthermore, an automatic reminder follow-up function also plays a pivotal  
7  
8 role during the whole follow-up period.  
9  
10

### 11 12 13 14 **Patient and public involvement**

15  
16 Patients and the public will not be involved in the development, design, conduct or  
17  
18 reporting of the study. The general results will be disseminated to participants through  
19  
20 public education during follow-up.  
21  
22  
23  
24  
25  
26

### 27 **Clinical outcomes**

28  
29 The primary study endpoint is postoperative recurrence following AF surgery defined  
30  
31 as the persistence or recurrence of AF symptoms, or the development of recurrent  
32  
33 perianal sepsis or chronic AF within 3 to 6 months of surgery.<sup>23 18</sup> The second end  
34  
35 point is a composite outcome of postoperative comorbidities or any equivalent events  
36  
37 including AFR, wound haemorrhage, faecal impaction, urinary retention, delayed  
38  
39 wound healing or unplanned hospitalization associated with AF surgery. Outcomes  
40  
41 were ascertained by the treating clinician combined with outpatient medical records or  
42  
43 patient self-reports.  
44  
45  
46  
47  
48  
49  
50  
51  
52

### 53 **Selection of predictor variables**

54  
55 Candidate variables for the prediction model of the composite outcome of  
56  
57 postoperative comorbidities will be screened according to the following pre-set  
58  
59  
60

1  
2  
3  
4 criteria: (1) prior clinical knowledge; (2) results from a systematic review updated in  
5  
6 November 2019 based on our published one<sup>4</sup> with sufficient evidence to include them  
7  
8 as predictive variables in the risk model for AFR as is demonstrated below; or (3)  
9  
10 agreed upon by a group of anorectal surgeons or experts for their clinical relevance  
11  
12 using a two-round Delphi survey. We initially identified the following covariates as  
13  
14 relevant candidate variables based on systematic reviews as well as clinical  
15  
16 knowledge and/or relevance. The determination of all other candidate variables is  
17  
18 based on the results of post-hoc analysis using univariable or multivariable survival  
19  
20 analyses with a threshold of  $p < 0.05$ .  
21  
22  
23  
24  
25  
26  
27  
28  
29

30 Factors identified from the systematic reviews and Delphi survey (manuscript under  
31  
32 review), will be measured at baseline. These include factors involving the identified  
33  
34 significant risk factors that are reported in our meta-analysis and are presented as  
35  
36 follows:  
37  
38

- 39 ▶ Prior anal surgery.
- 40
- 41 ▶ Seton placement surgery.
- 42
- 43 ▶ High transsphincteric fistula.
- 44
- 45 ▶ Internal opening unidentified.
- 46
- 47 ▶ Horseshoe extensions.
- 48
- 49 ▶ Multiple fistula tracts.
- 50
- 51
- 52
- 53
- 54
- 55
- 56
- 57
- 58
- 59
- 60

1  
2  
3  
4 Some of the demographic factors and surgery details will also be collected due to  
5  
6 limited power in the literature reviews and some non-significant potential factors (e.g.,  
7  
8 smoking or alcohol use) may be risk factors and are also included as follows:  
9  
10

- 11 ▶ Gender.
- 12
- 13 ▶ Age.
- 14
- 15
- 16 ▶ Smoking.
- 17
- 18
- 19 ▶ Alcohol use.
- 20
- 21
- 22 ▶ Diabetes mellitus.
- 23
- 24
- 25 ▶ Obesity.
- 26
- 27 ▶ Preoperative seton drainage.
- 28
- 29
- 30 ▶ High internal opening.
- 31
- 32 ▶ Postoperative drainage.
- 33
- 34
- 35 ▶ Supralevator extensions.
- 36

37 Data on other factors, such as laboratory examinations and MR imaging parameters  
38 (height of the internal openings, height and number of fistula, etc.) will also be  
39  
40 collected. Moreover, other factors, such as chronic steroid therapy, diverting stoma,  
41  
42 the surgeon's level of training, postoperative bowel confinement and antibiotic  
43  
44 prophylaxis reported in previous literature, were selected for regression analysis as  
45  
46 well. In addition, relevant factors from the expert-opinion survey were also assessed  
47  
48 including the number of prior anal fistula surgeries, the types of surgery performed  
49  
50 (such as staged fistulotomies, endorectal advancement flap and ligation of the  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

1  
2  
3  
4 intersphinteric fistula tract with and without seton drains), some nutrition parameters  
5  
6 and immunomodulation medication use.  
7  
8  
9

### 10 11 **Categorization of potential predictors**

12  
13  
14 We can code categorical predictors as “factor” variables, coding them as dummy  
15  
16 variables. For example, smoking is coded originally as “1” for never smoker, “2” for  
17  
18 past smoker, and “3” for current smoker and never smoker was selected as the  
19  
20 reference category. A similar approach can be used with alcohol use.<sup>19</sup> Continuous  
21  
22 variables formally should be measured with an interval or ratio scale, and should be  
23  
24 able to take any value in a range. We treat ordered variables as linear which is  
25  
26 generally reasonable for prediction. In other cases, continuous predictors can be  
27  
28 grouped with meaningful categorization; for example, body mass index (BMI) can be  
29  
30 classified based on internationally recognized categories (i.e., underweight, normal  
31  
32 weight, overweight, and obesity).<sup>20</sup> Based on previous experiences, we will derive  
33  
34 some predictors based on the responses of the surveys. However, in case some  
35  
36 subjectivity in the classifications of these predictors may occur, sensitivity analyses  
37  
38 will be performed to examine the robustness of our definitions during model  
39  
40 development and validation.  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52

### 53 **Study quality control for the prediction models**

54  
55  
56 Based on the summary of methodological quality and the developmental stage of  
57  
58 prediction models by van Oort et al., we are developing predesigned criteria for the  
59  
60

1  
2  
3  
4 quality control of our prediction models, which can allow us to conduct the study and  
5  
6 report the results more rigorously.<sup>21 22</sup> The methodological checklist of the study is  
7  
8 presented in the supplementary material.  
9  
10

### 11 12 13 14 **Missing data**

15  
16 Candidate predictors with more than 60% missingness will be excluded. For those  
17  
18 with less than 60% missingness, multiple imputation are to be performed by imputing  
19  
20 20 complete data sets using multivariate normal regression,<sup>23-26</sup> which can reasonably  
21  
22 approximate the true distributional relationship between the missing values and the  
23  
24 available ones.<sup>27</sup> Among various multiple imputation approaches, fully conditional  
25  
26 specification (FCS) and multivariate normal imputation (MVNI) are preferred,  
27  
28 because they have been proven to be generally less biased than complete-case analysis.  
29  
30 They can both generate similar results in the presence of either binary or ordinal  
31  
32 variables that are not generally normally distributed.<sup>27</sup>  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42

### 43 **Statistical analysis for model derivation**

44  
45 Logistic regression will be applied to develop our prediction models for the binary  
46  
47 outcomes. All data processing and statistical analysis will be performed using  
48  
49 EmpowerStats software (www. empowerstats.com; X&Y Solutions, Inc., Boston, MA,  
50  
51 USA) and the statistical software package R (R Foundation, Vienna, Austria).  
52  
53 We will first study the association between each potential variable and the outcome  
54  
55 based on univariable analysis. Variables are considered further for multivariable  
56  
57  
58  
59  
60

1  
2  
3  
4 regression modelling when they are associated with a p-value less than 0.20.  
5

6 Normality or linearity will be evaluated for the continuous predictors. Fractional  
7  
8  
9 polynomials are advocated for associations between the continuous predictors and the  
10  
11 outcome for nonlinear relationships.<sup>28 29</sup> We will perform backward stepwise  
12  
13 selection with  $p < 0.001$  as the inclusion threshold and  $p > 0.05$  as the exclusion  
14  
15 threshold for each imputed data set.  
16  
17  
18

19  
20  
21  
22 Predictors that appear in the imputation models with an inclusion fraction of  $\geq 50\%$   
23  
24 are qualified for the final multivariable model. Although there is no consensus  
25  
26 regarding the optimal method for selecting predictors for inclusion, backwards  
27  
28 elimination is generally considered as the preferred procedure as reported by Mantel  
29  
30 et al.<sup>30</sup> A forward stepwise procedure will also be performed to repeat the analysis to  
31  
32 test the robustness of the models. Overall regression coefficient estimates of the  
33  
34 models will be generated with the combination of the imputed datasets based on  
35  
36 Rubin's Rules, while taking into account uncertainty in the imputed values.<sup>23 24 26</sup>  
37  
38  
39  
40 Collinearity will also be assessed which refers to the fact that predictors can have  
41  
42 strong correlation with each other, defined as correlation coefficient  $> 0.8$ , or variance  
43  
44 inflation factor  $> 10$ .<sup>31</sup> Then we will examine the interactions among the regression  
45  
46  
47  
48  
49  
50  
51 models.  
52  
53  
54  
55

## 56 **Prediction model performance assessment**

57  
58  
59  
60

1  
2  
3  
4 Prediction models will be developed with a random sample of 60% of the AF cohort  
5  
6 as the derivation cohort, and then validated with the remaining sample of 40% of the  
7  
8 cohort as the validation cohort. The predictive performance in the derivation and  
9  
10 validation cohort will be evaluated and reported by examining measures of predictive  
11  
12 accuracy, discrimination and calibration. Nagelkerke's  $R^2$  and the Brier score will be  
13  
14 used for the measurement of predictive accuracy.<sup>32 33</sup> The discriminative ability of the  
15  
16 prediction models is evaluated using several statistics, which are according to the  
17  
18 discriminative and calibration ability in both the derivation and validation AF cohorts.  
19  
20 Model discrimination is the ability of the models to differentiate between high-risk  
21  
22 patients and low-risk patients (having high or low risk of AFR or surgery-related  
23  
24 complications). This will be assessed via Harrell's concordance statistic (C-index).<sup>34</sup>  
25  
26 The calculation of the C-index will be performed in each of the 20 imputed data sets,  
27  
28 and then averaged based on Rubin's rule.<sup>35</sup> The model is interpreted as having no  
29  
30 discriminatory ability when a value of C-index is 0.5, and has perfect discrimination  
31  
32 when a value of the C-index is 1.0.<sup>34</sup> Calibration implies the agreement between the  
33  
34 predicted outcomes and the observed outcomes, which is evaluated with the  
35  
36 Hosmer-Lemeshow goodness-of-fit test in all imputed datasets presented with  
37  
38 calibration plots.<sup>36</sup> Calibration-in-the-large, which is defined as the agreement  
39  
40 between mean observed outcomes and mean predictions, will also be assessed for  
41  
42 calibration.<sup>37</sup>  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

### **Internal and external validation of the prediction model**

1  
2  
3  
4 To make the prediction models reproducible, we must perform internal validation.  
5

6 The bootstrapping technique, as one of the most attractive resampling techniques, is a  
7  
8 mostly applied validation method that seems to be most efficient for obtaining stable  
9  
10 optimism-corrected estimates.<sup>34 38</sup> It has been reported that bootstrap validation is a  
11  
12 feasible technique for most prediction models with at least a 500 bootstrap resampling  
13  
14 procedures using Harrell's validation function, which can adjust the developed  
15  
16 models for overfitting.<sup>39</sup> We will also apply temporal validation as external validation  
17  
18 using a more recent AF patient cohort.<sup>40</sup>  
19  
20  
21  
22  
23  
24  
25  
26

### 27 **Sample size**

28  
29 Since there are no widely accepted methods for the estimation of the sample size  
30  
31 requirements to develop the risk prediction models, the size of this AF cohort will be  
32  
33 calculated to have 20 events per candidate predictive variable (EPV, defined as the  
34  
35 ratio of the number of individuals with the outcome event to the number of candidate  
36  
37 predictors), which can generally eliminate bias in regression coefficients for  
38  
39 prediction models with low-prevalence binary predictor development (the estimated  
40  
41 recurrence rate <20%) and adequately power the logistic regression models.<sup>41,42</sup>  
42  
43  
44  
45  
46  
47

48 According to the findings by Ogundimu et al.<sup>43</sup>, a higher EPV ( $\geq 20$ ) can generally  
49  
50 eliminate bias in regression coefficients for prediction models with low-prevalence  
51  
52 binary predictor development. Then we estimate 400 events allowing for 20 predictor  
53  
54 variables (EPV=20). Considering a 5-20% recurrence rate, we assume that at least  
55  
56 4000-8000 patients should be recruited for model development. In addition, surgery  
57  
58  
59  
60



1  
2  
3  
4 type and fistula type-stratified analyses will also be performed to examine the  
5  
6 different effects of these factors on disease recurrence or other related complications  
7  
8 in each subgroup. Risk prediction models, if possible, can also be developed in those  
9  
10 subpopulations. A cohort size with more than 4000-8000 patients will provide  
11  
12 sufficient power to perform those analyses and develop prediction models in those  
13  
14 subgroups.  
15  
16  
17  
18  
19  
20  
21

### 22 **Follow-up and methodological quality control**

23  
24 The application of WeChat questionnaires to collect data will inevitably increase the  
25  
26 probability of missing data. However, we have made some predesigned  
27  
28 countermeasures. For example, we have set up follow-up reminders via the WeChat  
29  
30 questionnaire system. Moreover, every week two trained clinical fellows cross-check  
31  
32 the data, and will contact the respondents by phone or WeChat about the missing  
33  
34 contents, which can minimize the missing data and loss to follow-up rate.  
35  
36  
37  
38  
39  
40  
41  
42

43 A limited number of studies have identified specific criteria for quality control in a  
44  
45 prediction model, but we have strictly adhered to the guidelines for the reporting of  
46  
47 developing studies, validating multivariable clinical prediction models as is reported  
48  
49 in the TRIPOD (Transparent Reporting of a multivariable prediction model for  
50  
51 Individual Prognosis Or Diagnosis) Statement to ensure methodological rigour.<sup>44</sup> All  
52  
53 issues have been addressed in this study in the supplementary material.  
54  
55  
56  
57  
58  
59  
60

## **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 recurrence 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,<sup>45-47</sup> lifestyle factors such as smoking and alcohol abuse;<sup>48 49</sup> 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<sup>48 50 51</sup>) are easily ascertainable before surgery may predict AFR. Similar risk prediction

1  
2  
3  
4 models exist in other diseases, such as the Framingham risk score model to predict  
5  
6 cardiovascular disease risk<sup>52</sup> and the Korean Crohn's disease prediction (KCDP)  
7  
8 model to predict the clinical course of Crohn's disease.<sup>53</sup> Until now, risk factor  
9  
10 investigations of predictors of perioperative surgery-related complications have been  
11  
12 limited to assessments of single predictors with small sample sizes. Risk prediction  
13  
14 models can help inform surgeons regarding high-risk AF patients based on the overall  
15  
16 risk factors. The primary purpose of this study was to develop two risk prediction  
17  
18 models to assist surgeons in identifying AF patients scheduled for surgical treatment  
19  
20 who are at higher risk of developing recurrence and surgery-related complications.  
21  
22 The predictive models will help both clinicians and patients identify the risk of  
23  
24 complications after AF surgery in advance, and perform the interventions necessary to  
25  
26 reduce the risk of surgery-related complications and the personal and social financial  
27  
28 burden brought about by those complications. Accurate risk prediction models are  
29  
30 especially instructive for the development of the optimal surgical plan to achieve  
31  
32 optimal surgical outcome.  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

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

1  
2  
3  
4 regression analysis. Second, we will apply the internal and external validation of the  
5  
6 prediction models using the bootstrapping procedure. Third, multiple imputation will  
7  
8 also be used to treat the missing data. Finally, our study is a prospective cohort study  
9  
10 with an adequate follow-up period which can minimize certain forms of bias.  
11  
12  
13  
14  
15  
16

17 Our study also has limitations. As many of the variables are collected through a  
18  
19 WeChat questionnaire system during hospitalization and follow-up, a higher  
20  
21 probability of missing data due to non-response bias may occur. To address this issue,  
22  
23 we regularly send reminders to those who do not respond after discharge. Second,  
24  
25 although we will investigate a series of potential predictors, some more potential  
26  
27 predictors will not be involved or not collected in the current study, such as data  
28  
29 related to postoperative nursing strategy and outpatient follow-up frequency.  
30  
31  
32  
33

34 Moreover, bias may also result from the single-centre recruitment of our study and  
35  
36 will be improved through multicentre recruitment in the future. Finally, one key  
37  
38 potential issue that needs to be considered a priori is the variable recurrence rate  
39  
40 depending on the risk factors included and identified in the testing and validating  
41  
42 cohorts since this may affect the C-index. It is possible that a lack of key variable  
43  
44 inclusions in the models may result in decreased discriminatory ability. This is a  
45  
46 function of the database and points of interest included that may need to be  
47  
48 maximized before proceeding with development of the testing model.  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

1  
2  
3  
4 The newly developed risk algorithms may have significant applications in clinical  
5  
6 practice by helping recommend an optimal surgical approach for a specific AF patient,  
7  
8 as well as intensive perioperative care, education, and the timely assessment and  
9  
10 discussion of the need for interventions among those most at the highest risk of  
11  
12 developing recurrence or surgery-related complications. The models will specifically  
13  
14 identify AF patients who are likely to develop recurrence or related complications  
15  
16 following AF surgery to offer a quantitative evaluation of the risk. Moreover, the  
17  
18 models will also provide reference information for preventing recurrence, reducing  
19  
20 the rate of recurrence after operation and intervening in high-risk factors in the early  
21  
22 stage.  
23  
24  
25  
26  
27  
28  
29  
30  
31

32 In summary, this study protocol summarizes the design of development and validation  
33  
34 studies for a risk screening tool in patients receiving AF surgery. The results from this  
35  
36 study will be interpreted for the purpose of clinical decision making. The models to be  
37  
38 developed in this study could be used to make new recommendations for  
39  
40 perioperative AF patients.  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

## References

1. Tantiplachiva 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.ijssu.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

- 1  
2  
3  
4 7. Gottgens KW, Janssen PT, Heemskerk J, et al. Long-term outcome of low perianal  
5  
6 fistulas treated by fistulotomy: a multicenter study. *International journal of*  
7  
8 *colorectal disease* 2015;30(2):213-9. doi: 10.1007/s00384-014-2072-y  
9  
10  
11  
12 8. Bakhtawar N, Usman M. Factors Increasing the Risk of Recurrence in Fistula-in-ano.  
13  
14 *Cureus* 2019;11(3):e4200. doi: 10.7759/cureus.4200  
15  
16  
17  
18 9. Sygut A, Mik M, Trzcinski R, et al. How the location of the internal opening of anal  
19  
20 fistulas affect the treatment results of primary transsphincteric fistulas.  
21  
22 *Langenbeck's archives of surgery* 2010;395(8):1055-9. doi:  
23  
24 10.1007/s00423-009-0562-0  
25  
26  
27  
28  
29  
30 10. Dziki A, Bartos M. Seton treatment of anal fistula: experience with a new  
31  
32 modification. *The European journal of surgery = Acta chirurgica* 1998;164(7):543-8.  
33  
34 doi: 10.1080/110241598750005930  
35  
36  
37  
38  
39 11. Hermann J, Eder P, Banasiewicz T, et al. Current management of anal fistulas in  
40  
41 Crohn's disease. *Przegląd gastroenterologiczny* 2015;10(2):83-8. doi:  
42  
43 10.5114/pg.2015.49684  
44  
45  
46  
47  
48 12. Hamadani A, Haigh PI, Liu IL, et al. Who is at risk for developing chronic anal  
49  
50 fistula or recurrent anal sepsis after initial perianal abscess? *Diseases of the colon*  
51  
52 *and rectum* 2009;52(2):217-21. doi: 10.1007/DCR.0b013e31819a5c52  
53  
54  
55  
56  
57  
58  
59  
60

- 1  
2  
3  
4 13. Ellis CN, Clark S. Effect of tobacco smoking on advancement flap repair of  
5  
6 complex anal fistulas. *Diseases of the colon and rectum* 2007;50(4):459-63. doi:  
7  
8 10.1007/s10350-006-0829-2  
9  
10  
11  
12 14. Moons KG, Altman DG, Vergouwe Y, et al. Prognosis and prognostic research:  
13  
14 application and impact of prognostic models in clinical practice. *Bmj* 2009;338:b606.  
15  
16 doi: 10.1136/bmj.b606  
17  
18  
19  
20 15. Ming C, Viassolo V, Probst-Hensch N, et al. Machine learning techniques for  
21  
22 personalized breast cancer risk prediction: comparison with the BCRAT and  
23  
24 BOADICEA models. *Breast cancer research : BCR* 2019;21(1):75. doi:  
25  
26 10.1186/s13058-019-1158-4  
27  
28  
29  
30  
31  
32 16. Khan SS, Ning H, Shah SJ, et al. 10-Year Risk Equations for Incident Heart Failure  
33  
34 in the General Population. *Journal of the American College of Cardiology*  
35  
36 2019;73(19):2388-97. doi: 10.1016/j.jacc.2019.02.057  
37  
38  
39  
40  
41 17. Terry MB, Liao Y, Whittemore AS, et al. 10-year performance of four models of  
42  
43 breast cancer risk: a validation study. *The Lancet Oncology* 2019;20(4):504-17. doi:  
44  
45 10.1016/S1470-2045(18)30902-1  
46  
47  
48  
49 18. Yano T, Asano M, Matsuda Y, et al. Prognostic factors for recurrence following  
50  
51 the initial drainage of an anorectal abscess. *International journal of colorectal*  
52  
53 *disease* 2010;25(12):1495-8. doi: 10.1007/s00384-010-1011-9 [published Online  
54  
55 First: 2010/07/20]  
56  
57  
58  
59  
60



1  
2  
3  
4 19. Steyerberg EW. Coding of Categorical and Continuous Predictors. In: Steyerberg  
5  
6 EW, ed. *Clinical Prediction Models: A Practical Approach to Development, Validation,*  
7  
8 *and Updating.* Cham: Springer International Publishing 2019:175-90.

9  
10  
11  
12 20. Morgan TM, Elashoff RM. Effect of Categorizing a Continuous Covariate on the  
13  
14  
15 Comparison of Survival Time. *Journal of the American Statistical Association*  
16  
17  
18 1986;81(396):917-21. doi: 10.1080/01621459.1986.10478350

19  
20  
21 21. van Oort L, van den Berg T, Koes BW, et al. Preliminary state of development of  
22  
23  
24 prediction models for primary care physical therapy: a systematic review. *Journal of*  
25  
26  
27 *Clinical Epidemiology* 2012;65(12):1257-66. doi:  
28  
29 <https://doi.org/10.1016/j.jclinepi.2012.05.007>

30  
31  
32  
33 22. Rushton AB, Verra ML, Emms A, et al. Development and validation of two clinical  
34  
35  
36 prediction models to inform clinical decision-making for lumbar spinal fusion surgery  
37  
38  
39 for degenerative disorders and rehabilitation following surgery: protocol for a  
40  
41  
42 prospective observational study. *BMJ Open* 2018;8(5):e021078. doi:  
43  
44  
45 [10.1136/bmjopen-2017-021078](https://doi.org/10.1136/bmjopen-2017-021078)

46  
47 23. Wood AM, White IR, Royston P. How should variable selection be performed  
48  
49  
50 with multiply imputed data? *Statistics in Medicine* 2008;27(17):3227-46. doi:  
51  
52  
53 [10.1002/sim.3177](https://doi.org/10.1002/sim.3177)

- 1  
2  
3  
4 24. White IR, Royston P, Wood AM. Multiple imputation using chained equations:  
5  
6 Issues and guidance for practice. *Statistics in Medicine* 2011;30(4):377-99. doi:  
7  
8 10.1002/sim.4067  
9  
10  
11  
12 25. Schafer JL. Multiple imputation: a primer. *Statistical Methods in Medical*  
13  
14 *Research* 1999;8(1):3-15. doi: 10.1177/096228029900800102  
15  
16  
17  
18 26. Marshall A, Altman DG, Holder RL, et al. Combining estimates of interest in  
19  
20 prognostic modelling studies after multiple imputation: current practice and  
21  
22 guidelines. *BMC Medical Research Methodology* 2009;9(1):57. doi:  
23  
24 10.1186/1471-2288-9-57  
25  
26  
27  
28  
29 27. Lee KJ, Carlin JB. Multiple Imputation for Missing Data: Fully Conditional  
30  
31 Specification Versus Multivariate Normal Imputation. *American Journal of*  
32  
33 *Epidemiology* 2010;171(5):624-32. doi: 10.1093/aje/kwp425  
34  
35  
36  
37  
38 28. Sauerbrei W, Royston P, Binder H. Selection of important variables and  
39  
40 determination of functional form for continuous predictors in multivariable model  
41  
42 building. *Statistics in Medicine* 2007;26(30):5512-28. doi: 10.1002/sim.3148  
43  
44  
45  
46  
47 29. Royston P, Sauerbrei W. Building multivariable regression models with  
48  
49 continuous covariates in clinical epidemiology--with an emphasis on fractional  
50  
51 polynomials. *Methods of information in medicine* 2005;44(4):561-71.  
52  
53  
54  
55  
56 30. Mantel N. Why Stepdown Procedures in Variable Selection. *Technometrics*  
57  
58 1970;12(3):621-25. doi: 10.1080/00401706.1970.10488701  
59  
60

1  
2  
3  
4 31. Vittinghoff E, Glidden DV, Shiboski SC, et al. Regression methods in biostatistics:  
5  
6 linear, logistic, survival, and repeated measures models: Springer Science & Business  
7  
8 Media 2011.  
9

10  
11  
12 32. Steyerberg EW, Vickers AJ, Cook NR, et al. Assessing the performance of  
13  
14 prediction models: a framework for traditional and novel measures. *Epidemiology*  
15  
16 2010;21(1):128-38. doi: 10.1097/EDE.0b013e3181c30fb2  
17  
18  
19

20  
21 33. NAGELKERKE NJD. A note on a general definition of the coefficient of  
22  
23 determination. *Biometrika* 1991;78(3):691-92. doi: 10.1093/biomet/78.3.691  
24  
25  
26

27  
28 34. Harrell Jr FE. Regression modeling strategies: with applications to linear models,  
29  
30 logistic and ordinal regression, and survival analysis: Springer 2015.  
31  
32

33  
34 35. Rubin DB. Multiple imputation for nonresponse in surveys: John Wiley & Sons  
35  
36 2004.  
37  
38

39  
40 36. Hosmer Jr DW, Lemeshow S, Sturdivant RX. Applied logistic regression: John  
41  
42 Wiley & Sons 2013.  
43  
44

45  
46 37. Snell KI, Ensor J, Debray TP, et al. Meta-analysis of prediction model performance  
47  
48 across multiple studies: Which scale helps ensure between-study normality for the  
49  
50 C-statistic and calibration measures? *Statistical methods in medical research*  
51  
52 2018;27(11):3505-22.  
53  
54  
55  
56  
57  
58  
59  
60

- 1  
2  
3  
4 38. Chatfield C. Model uncertainty, data mining and statistical inference. *Journal of*  
5  
6 the Royal Statistical Society: Series A (Statistics in Society) 1995;158(3):419-44.  
7  
8  
9  
10 39. Steyerberg EW, Bleeker SE, Moll HA, et al. Internal and external validation of  
11  
12 predictive models: a simulation study of bias and precision in small samples. *Journal*  
13  
14 of clinical epidemiology 2003;56(5):441-47.  
15  
16  
17  
18 40. Justice AC, Covinsky KE, Berlin JA. Assessing the generalizability of prognostic  
19  
20 information. *Annals of internal medicine* 1999;130(6):515-24.  
21  
22  
23  
24 41. Concato J, Peduzzi P, Holford TR, et al. Importance of events per independent  
25  
26 variable in proportional hazards analysis I. Background, goals, and general strategy.  
27  
28 *Journal of clinical epidemiology* 1995;48(12):1495-501.  
29  
30  
31  
32  
33 42. Steyerberg EW. *Clinical Prediction Models: A Practical Approach to Development,*  
34  
35 *Validation, and Updating.* SPRINGER NATURE, 2019.  
36  
37  
38  
39 43. Ogundimu EO, Altman DG, Collins GS. Adequate sample size for developing  
40  
41 prediction models is not simply related to events per variable. *J Clin Epidemiol.*  
42  
43 2016;76:175–182.  
44  
45  
46  
47  
48 44. Moons KG, Altman DG, Reitsma JB, et al. Transparent Reporting of a  
49  
50 multivariable prediction model for Individual Prognosis or Diagnosis (TRIPOD):  
51  
52 explanation and elaboration. *Annals of internal medicine* 2015;162(1):W1-W73.  
53  
54  
55  
56  
57  
58  
59  
60

- 1  
2  
3  
4 45. Hamadani A, Haigh PI, Liu I-LA, et al. Who is at risk for developing chronic anal  
5  
6 fistula or recurrent anal sepsis after initial perianal abscess? *Diseases of the colon &*  
7  
8 *rectum* 2009;52(2):217-21.  
9  
10  
11  
12 46. Sonoda T, Hull T, Piedmonte MR, et al. Outcomes of primary repair of anorectal  
13  
14 and rectovaginal fistulas using the endorectal advancement flap. *Diseases of the*  
15  
16 *colon & rectum* 2002;45(12):1622-28.  
17  
18  
19  
20 47. Hyman N, O'Brien S, Osler T. Outcomes after fistulotomy: results of a prospective,  
21  
22 multicenter regional study. *Diseases of the Colon & Rectum* 2009;52(12):2022-27.  
23  
24  
25  
26 48. Ellis CN, Clark S. Effect of tobacco smoking on advancement flap repair of  
27  
28 complex anal fistulas. *Diseases of the colon & rectum* 2007;50(4):459-63.  
29  
30  
31  
32 49. Schwandner T, Roblick MH, Kierer W, et al. Surgical treatment of complex anal  
33  
34 fistulas with the anal fistula plug: a prospective, multicenter study. *Diseases of the*  
35  
36 *Colon & Rectum* 2009;52(9):1578-83.  
37  
38  
39  
40  
41  
42 50. Garcia-Aguilar J, Belmonte C, Wong WD, et al. Anal fistula surgery: Factors  
43  
44 associated with recurrence and incontinence. *Diseases of the colon & rectum*  
45  
46 1996;39(7):723-29.  
47  
48  
49  
50  
51 51. Athanasiadis S, Köhler A, Nafe M. Treatment of high anal fistulae by primary  
52  
53 occlusion of the internal ostium, drainage of the intersphincteric space, and mucosal  
54  
55 advancement flap. *International journal of colorectal disease* 1994;9(3):153-57.  
56  
57  
58  
59  
60

1  
2  
3  
4 52. Fox ER, Samdarshi TE, Musani SK, et al. Development and validation of risk  
5  
6 prediction models for cardiovascular events in black adults: the Jackson Heart Study  
7  
8 Cohort. JAMA cardiology 2016;1(1):15-25.  
9  
10

11  
12 53. Park Y, Cheon JH, Park YL, et al. Development of a novel predictive model for the  
13  
14 clinical course of Crohn's disease: results from the connect study. Inflammatory  
15  
16 bowel diseases 2017;23(7):1071-79.  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

1  
2  
3 **Figure legend**  
4  
5

6 Figure 1. Flowchart of prediction model development and assessment.  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

For peer review only

Figure 1: Flowchart of prediction model development and assessment.

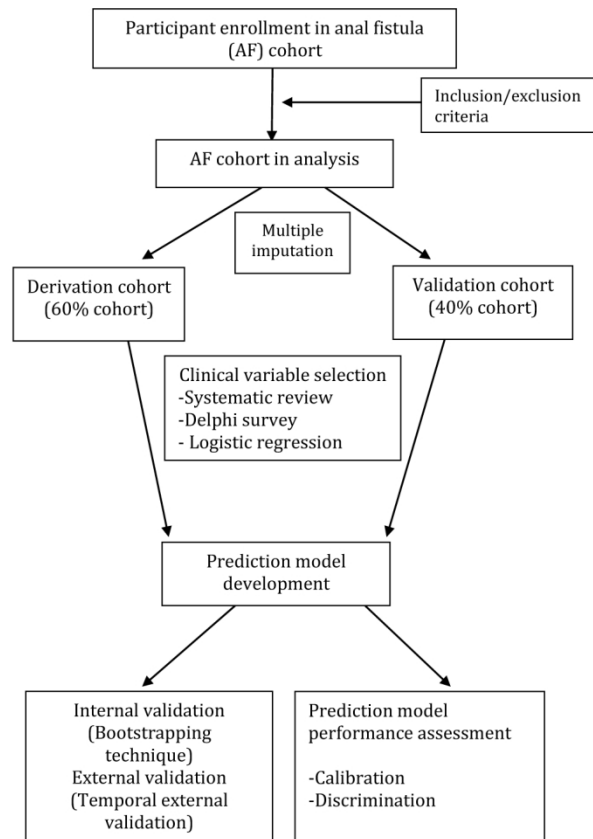


Figure 1. Flowchart of prediction model development and assessment.





## 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
<b>Title and abstract</b>				
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
<b>Introduction</b>				
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
<b>Methods</b>				
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
<b>Results</b>				
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	-

## TRIPOD Checklist: Prediction Model Development and Validation

specification	15b	D	coefficients, and model intercept or baseline survival at a given time point). Explain how to 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).	-
<b>Discussion</b>				
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
<b>Other information</b>				
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.

# 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-ECA7-4F33-393A-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](https://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](mailto: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](https://authorservices.springernature.com).

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

1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47

## 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
<b>Title and abstract</b>				
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
<b>Introduction</b>				
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
<b>Methods</b>				
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
<b>Results</b>				
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	-



## 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).	-
<b>Discussion</b>				
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
<b>Other information</b>				
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.