BMJ Open

Validating traditional Chinese syndrome features in varied stages of chronic gastritis malignant transformation: study protocol for a cross-sectional study

Journal:	*BMJ Open*
Manuscript ID	bmjopen-2017-020939
Article Type:	Protocol
Date Submitted by the Author:	01-Dec-2017
Complete List of Authors:	Zhang, Yin; Bregham and Women's Hospital and Harvard Medical School, Medicine; Beijing University of Chinese Medicine, School of Chinese Medicine Liu, Yue; Beijing University of Chinese Medicine, School of Chinese Medicine Song, Rui; The Association of Chinese American Physicians; American Chinese Medical Exchange Society Zhang, Li; China-Japan Friendship Hospital, Department of Neurosurgery Su, Zeqi; Beijing University of Chinese Medicine, School of Chinese Medicine Li, Yannan; Affiliated Hospital of Shandong University of Chinese Medicine, Department of Gynecology Chen, Runhua; Dongfang Hospital Affiliated to Beijing University of Chinese Medicine, Department of Gastroenterology Shi, Ning; Beijing Electric Power Hospital, Department of Health Management Zhao, Xia; Beijing University of Chinese Medicine, School of Humanities Du, Shiyu; China-Japan Friendship Hospital, Department of Gastroenterology Ding, Xia; Beijing University of Chinese Medicine, School of Chinese Medicine
Keywords:	Chronic gastritis, Malignant transformation, Traditional Chinese medicine, Syndrome, Cross-sectional study

Validating traditional Chinese syndrome features in varied stages of chronic gastritis malignant transformation: study protocol for a cross-sectional study

Yin Zhang,[1,2,3,a] Yue Liu,[1,a] Rui Song,[4,5,a] Li Zhang,[6] Zeqi Su,[1] Yannan Li,[7] Runhua Chen,[8,1] Ning Shi,[9,1] Xia Zhao,[10] Shiyu Du,[11] and Xia Ding [1,*]

[1]School of Chinese Medicine, Beijing University of Chinese Medicine, Chaoyang District, Beijing 100029, China

[2]Channing Division of Network Medicine, Department of Medicine, Brigham and Women's Hospital and Harvard Medical School, Boston, MA 02115, USA

[3]Institute of Basic Research in Clinical Medicine, China Academy of Chinese Medical Sciences, Dongcheng District, Beijing 100700, China

[4]The Association of Chinese American Physicians, Flushing, NY 11354, USA

[5]American Chinese Medical Exchange Society, Burlington, MA 01803, USA

[6]Department of Neurosurgery, China-Japan Friendship Hospital, Chaoyang District, Beijing 100029, China

[7]Department of Gynecology, Affiliated Hospital of Shandong University of Chinese Medicine, Lixia District, Jinan 250011, China

[8]Department of Gastroenterology, Dongfang Hospital Affiliated to Beijing University of Chinese Medicine, Beijing 100078, China

[9]Department of Health Management, Beijing Electric Power Hospital, Fengtai District, Beijing 100073, China

[10]School of Humanities, Beijing University of Chinese Medicine, Chaoyang District, Beijing 100029, China

[11]Department of Gastroenterology, China-Japan Friendship Hospital, Chaoyang District, Beijing 100029, China

[a] These authors contributed equally to this work.

[*] Correspondence to: Dr. Xia Ding, E-mail: dingx@bucm.edu.cn

Word count: 2900 words (excluding title page, abstract, references, figures and tables)

Abstract

Introduction

The transition from chronic non-atrophic gastritis (CNAG) to chronic atrophic gastritis (CAG) and gastric carcinoma (GC) is regarded as a representative disease model of gastric mucosa malignant transformation leaded by uncontrolled inflammation. Traditional Chinese medicine (TCM) syndrome-targeted therapies have been applied in treating CG (chronic gastritis) malignant transformation in China with satisfying efficacy. This study aims to validate TCM syndrome features in each stage of CG malignant transformation. The findings may shed light on the TCM pathogenesis theory of CG malignant transformation, and thus provides potential translation for optimization of syndrome-targeted treatment strategies of CNAG, CAG, and GC respectively.

Methods and analysis

The present study is a cross-sectional study conducted in China. 2000 eligible patients, including 500 CNAG cases, 1000 CAG cases, and 500 GC cases, will be recruited from 4 TCM hospitals. Descriptive analysis, comparative analysis, and correlation analysis of all the measurement data will be performed by biostatisticians. Unsupervised data mining analyses, including exploratory factor analysis, association rule analysis, hierarchical clustering analysis, and complex system entropy clustering analysis, etc., will also be performed by data scientists respectively for indepth analyses of TCM syndrome-related indicators.

Ethics and dissemination

The protocol has been approved by Ethical Review Board of Donzhimen Hospital Affiliated to Beijing University of Chinese Medicine (No. ECPJ-BDY-2014-02). All the study outcomes will be disseminated through national conference reports and meantime published on peer-reviewed journals.

Study registration

This study has been registered on ClinicalTrials.gov (No. NCT03314038).

Keywords:

Chronic gastritis; Malignant transformation; Traditional Chinese medicine; Syndrome; Cross-sectional study

Strengths and limitations of this study

►To the authors' knowledge, this is among the first reported large population based studies on validating TCM syndrome features in varied stages of CG malignant transformation.

►A large sample size of 2000 participants can be obtained. Full-scale TCM and modern medicine indicators can be accurately collected.

►Both statistical analysis and unsupervised data mining strategies will be applied by research personnel parallelly to reinforce complementary advantages.

►The findings may shed light on optimization of both TCM syndrome-related pathogenesis theory and syndrome-targeted interventions of CG malignant transformation.

►This is not a prospective study, and thus longitudinal data couldn't be collected by research personnel to speculate causal relationship.

►Selection bias and information bias may exist.

Introduction

Chronic gastritis (CG), persistent inflammation of the gastric mucosa with no specific clinical manifestations, has been regarded as the most common gastrointestinal disease with an incidence of 70% among the adult population in China [1]. Chronicity and recurrence of chronic non-atrophic gastritis (CNAG) may lead to chronic atrophic gastritis (CAG), which is characterized by gastric mucosal atrophy and usually accompany with metaplasia or dysplasia [2]. As the most important intermediate step of CG malignant transformation, CAG is acknowledged as premalignant lesion of gastric cancer (GC). The sequential stages of inflammation, metaplasia, dysplasia and carcinoma have been proved a well-established tumorigenesis model of intestinal type of GC leaded by uncontrolled inflammation [2-4]. Statistics have shown that there are 951,600 annual new cases of GC worldwide [5]. China has been reported as one of the countries with largest population of both CG and GC sufferers [1]. Considering the severe burden of above-mentioned diseases, active intervention on CG malignant transformation process is highly needed [6-8].

Traditional Chinese medicine (TCM), as one of the most important and time-honored alternative medicine approaches worldwide, has developed unique theories of etiology and diagnosis since ancient time. TCM practitioners have been applying the central principles of syndrome pattern differentiation in clinical practice to identify the physical condition of patients and the pathologensis of diseases with a history of more than 2500 years. Nowadays, TCM physicians keep on using the prestigious and traditional "four diagnostic methods" of looking, listening/smelling, asking and palpating to collect the information of symptoms and signs on patients comprehensively. Full-scale assessment of aforementioned clinical information can then be conducted by physicians based on TCM theory and clinical experience to reveal the etiology of diseases, and thus guiding individualized herbal prescriptions. Though huge leaps have been seen in past decades in elaborating the pathogenesis of CAG, however, modern medicines in treating CNAG and CAG remain unsatisfied to some extent [9]. Owing to the heavy burden of recurrence, chronicity, and malignant transformation of CG, a considerable proportion of sufferers in China have put their concentrations on TCM therapies.

Trials have indicated the benefits of TCM syndrome-targeted interventions on CNAG and CAG [10-14]. The Society of Gastroenterology, China Association of Chinese Medicine has propagated the latest "*TCM Consensus on Chronic Superficial Gastritis Diagnosis and Treatment*" and "*TCM Consensus on Chronic Atrophic Gastritis Diagnosis and Treatment*" respectively in 2009 [15, 16]. The Society of Gastroenterology, China Society of Integrated Traditional Chinese and Western Medicine also issued "*Integrative Medicine Consensus on Chronic Gastritis Diagnosis and Treatment in China*" in 2012 [17]. The aforementioned consensuses emphasized the importance of applying TCM syndrome-targeted therapies in controlling gastric mucosal inflammation and atrophy, and inhibiting metaplasia and dysplasia in accordance with China's local conditions. "*Consensus on chronic gastritis in China*" promulgated by the Society of Gastroenterology, Chinese Medical Association and "*Diagnosis and Treatment Guideline for Gastric Carcinoma*" issued by the National Health and Family Planning Commission of the People's Republic of China also recommended the application of TCM syndrome differentiation based therapies in addition to conventional treatment respectively [1, 18].

Both consensus and literature reviews have also indicated the potential links between syndrome patterns and CG malignant transformation [15-17, 19-21]. We believe that the diversity and dynamics of TCM syndromes may play a role all the way through this course. However, no large population based study

has been reported to support in-depth analysis of prevalence and severity of varied TCM syndromes in different stages of CG malignant transformation. Taking note of the limitations of previous studies, we highlight validating detailed characteristics of TCM syndromes in CNAG, CAG, and GC population respectively in the present study. The new findings may shed light on the TCM pathogenesis theory of CG malignant transformation, and thus provide potential translation for optimization of syndrome-targeted treatment strategies in varied stages of this process.

Study aims

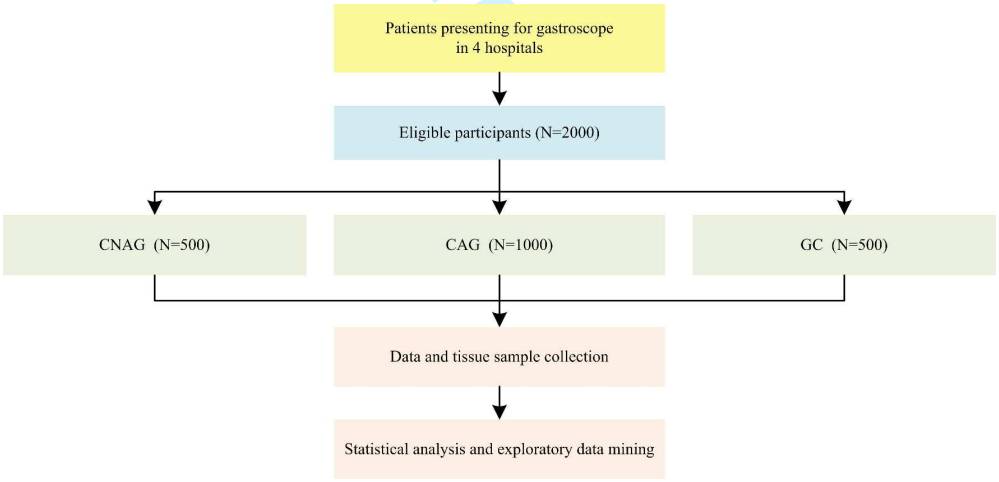
The primary motivation for the present study is to identify TCM syndrome features in varied stages of CG malignant transformation respectively, including CNAG, CAG, and GC. In addition, this study can also be regarded as a pilot study speculating potential dynamic features of TCM syndrome in the whole course of CG. This may provide important and targeted clues for the design of future observational long-term follow-up studies aiming to further explore novel risk prediction and syndrome evolution models combining both TCM and modern medicine indicators for CG or CAG malignant transformation.

Methods

Design

This is a cross-sectional study undertaken from 2014 to 2018, comprising 2000 eligible participants in 4 hospitals in China. The general flow diagram of this study is outlined in Fig. 1.

Fig.1 General flow diagram of this study



Recruitment

This study is leaded by Beijing University of Chinese Medicine (BUCM). The 4 medical centres participating in the present study include Dongzhimen Hospital Affiliated to BUCM, Dongfang Hospital Affiliated to BUCM, the 3rd Affiliated Hospital of BUCM, and Wangjing Hospital Affiliated to China Academy of Chinese Medical Sciences (CACMS). Taking into full consideration of items of CRFs, experts' opinions, study expenses and operability, the anticipated sample size is decided to be 2000 in total, including 500 CNAG cases, 1000 CAG cases, and 500 GC cases. An upper gastrointestinal endoscopy based screening process is conducted by study personnel during recruitment. All the eligible patients presenting to the the aforementioned hospitals are then identified and recruited continuously until the

required sample size for each category (CNAG, CAG, and GC) is achieved respectively. This study has opened to recruitment in 2014. Recruitment is ongoing and is expected to conclude by the end of 2018.

Diagnostic Criteria

Diagnostic criteria of CNAG, CAG, GMA, IM, and GED in the present study refers to "Consensus on chronic gastritis in China" promulgated by the Society of Gastroenterology, Chinese Medical Association [1]. Diagnostic criteria of GC refers to "*Diagnosis and Treatment Guideline for Gastric Carcinoma (trail version)*" issued by the National Health and Family Planning Commission of the People's Republic of China [18]. Diagnostic criteria of Helicobacter pylori (Hp) infection refers to "*Chinese Consensus on Helicobacter pylori Infection Treatment*" issued by the Chinese Society of Gastroenterology, Chinese Medical Association [22]. Criteria for TCM syndrome pattern differentiation diagnosis in this study refers to "*TCM Consensus on Chronic Superficial Gastritis Diagnosis and Treatment*" [15] and "*TCM Consensus on Chronic Atrophic Gastritis Diagnosis and Treatment*" [16] issued by the Society of Gastroenterology, China Association of Chinese Medicine.

Inclusion Criteria

The inclusion criteria are as follows: (1) Meeting diagnostic criteria of CNAG, CAG, or GC; (2) Willing to cooperate with investigators for data and tissue sample collection; (3) Willing to sign informed consent.

Exclusion Criteria

The exclusion criteria are as follows: (1) History of previous stomach surgeries; (2) Unable to participate in data and sample collection for any reason.

Data and samples collection

Strict training in standard operating procedure (SOP) of clinical data and samples collection is conducted by principal investigator for research personnel before study initiation. Prior to recruitment, participants are required to sign informed consents. In this process, clinial investigators take the responsibility for explaining the objectives, general procedures, data collection methods, risks and benefits, authorization, data privacy policy, and other necessary details to every patient. Research personnel will then assess all the indicators and tissue samples needed at once in recruitment. Each participant is interviewed by at least two physicians to ensure quality control. A list of degree-definition of all the TCM indicators is attached to each CRF for investigators and participants to refer to.

Outcome measures and overview of CRFs

Primary measures of this study include the prevalence of TCM syndrome patterns in varied stages of CG malignant transformation (CNAG, CAG, and GC). Secondary outcome measures include prevalence and severity of all the presenting signs and symptoms collected by using TCM four diagnostic methods (looking, listening/smelling, asking, and palpating). The aforementioned clinical features are identified in detail base on TCM syndrome pattern differentiation scales.

Accurate modern medicine diagnostic information in this study is achieved based on latest gastroscope reports, and pathological reports. Hp infects more than half of the populations worldwide and causes long-term progressive damage to the gastric mucosa in all infected individuals, is acknowledged as one of

the most important risk factors for CG and GC [23-26]. Stage-specific dietary intake, ingestion of ascorbic acid and nitrate, and other factors have also been proved to associate with the multistep and multifactorial process of gastric carcinogenesis [1-4, 27]. Considering the importance of above-mentioned indicators, full-scale information regarding to Hp test reports, dietary intake, life behaviors, and medication, etc., are collected in addition to TCM indicators. Other necessary information collected in our study covers demographic information, disease history, complications, and laboratory indicators. In addition, tissue samples collection is also included. Subgroup analysis can thus be carried out in a number of specific populations (e.g. CAG with metaplasia and/or dysplasia, CAG with/without Hp infection, etc.), which is of significant clinical value.

According to study design and case report forms (CRFs), detailed content of data and tissue samples collection are illustrated in Table. 1.

Table.1 Content of data and tissue samples collection

Content	Details
Demographic Information	Name, date of birth, gender, age, nationality, height, weight, level of education, etc.
Diagnostic Information	Latest gastroscope reports, Hp test reports, and pathological reports
TCM Syndrome Related Information	Features of presenting signs and symptoms collected by using TCM four diagnostic methods of looking, listening/smelling, asking, and palpating
	TCM syndrome pattern differentiation scale scores
	Results of TCM syndrome identification
Disease History	Past history and family history of CG, GC, and Hp infection, etc.
Complications	Upper gastrointestinal hemorrhage, upper gastrointestinal ulcer, pyloric obstruction, anemia, gastric carcinoma metastasis, etc.
Dietary Intake and Life Behaviors	Dietary structure, working state, psychological state, alcohol intake history, history of smoking, etc.
Medication	Current and past drug use on CG and GC treatment, and nonsteroidal anti-inflammatory drugs (NSAIDs)
Laboratory Indicators	Routine blood test, routine stool test, tumor markers, etc.
Tissue Samples	Tissue samples of diseased gastric mucosa

Data management

SOP of data collection, entry, editing, locking and retrieval is set up by data management centre of this study in Beijing University of Chinese Medicine. Confidentiality, authenticity, and integrity of all the clinical information gathered in the present study is highlighted and maintained at all levels of data management. CRFs are organized by unique identification numbers and thus no patients' personal identifiers will be available in any records. Before data-entry, double-checking of range and logic of every

variable value is performed by primary researchers. Data is then double-entered into a dedicated database by data administrators to confirm there is no input error.

Data analysis and outcome reporting

In order to achieve the primary goal of validating TCM syndrome features in varied stages of CG malignant transformation, varied data analysis strategies on multidimensional indicators will be performed parallelly from both biostatistics and data mining perspectives.

Descriptive statistics will be carried out by biostatisticians for all the measurement data, including demographic and clinical characteristics in the whole study population and each stratum. Comparative analysis of qualitative TCM and modern medicine indicators in different groups will be performed using chi-squared tests. Comparative analysis of quantitative scores of TCM syndrome pattern differentiation scales will be conducted using rank-sum test. A probability of $P < 0.05$ will be considered statistically significant for these analyses. Correlation analysis of pathological grades and stages of CG malignant transformation and TCM indicators will be performed using logistic regression model. The significance level for introducing and removing variables will be 0.05 and 0.10.

In addition, unsupervised data mining analyses, including exploratory factor analysis, association rule analysis, hierarchical clustering analysis, and complex system entropy clustering analysis, etc., will be performed by data scientists respectively for indepth exploring of TCM syndrome distribution and evolution features in the whole course of CG malignant transformation.

Statistical analyses in this study will be carried out using SAS software *(version 9.3, SAS Institute Inc., Cary, NC, U.S.A)* , and R software *(version 3.3.1, The R Project for Statistical Computing., Auckland, N.Z.)*. Data mining analyses in this study will be processed by SPSS Clementine software *(version 12.0, SPSS Inc., Chicago, IL, U.S.A)*, and Matlab Software *(version 2016a, Mathworks Inc., Natick, MA, U.S.A)*. Plotting in this study will be performed using Office Excel software *(version 2007, Microsoft Inc., Redmond, WA, U.S.A)* and Office Visio software *(version 2007, Microsoft Inc., Redmond, WA, U.S.A)*.

General flow diagram of biostatistical ananlysis process is outlined in Fig. 2. Detailed flow diagrams of aforementioned exploratory data mining strategies to be conducted in this study are illustrated respectively in Fig. 3-5.

Fig.2 Flow diagram of statistical analysis strategy in this study

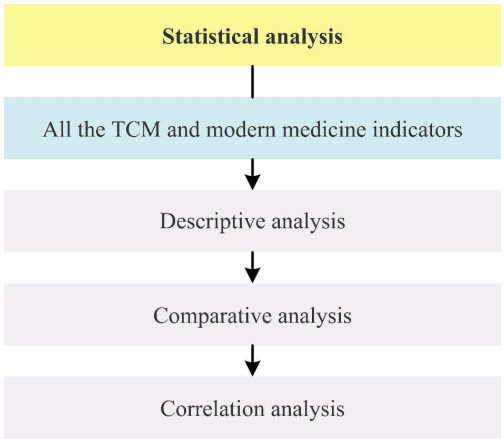


Fig.3 Flow diagram of exploratory factor analysis in this study

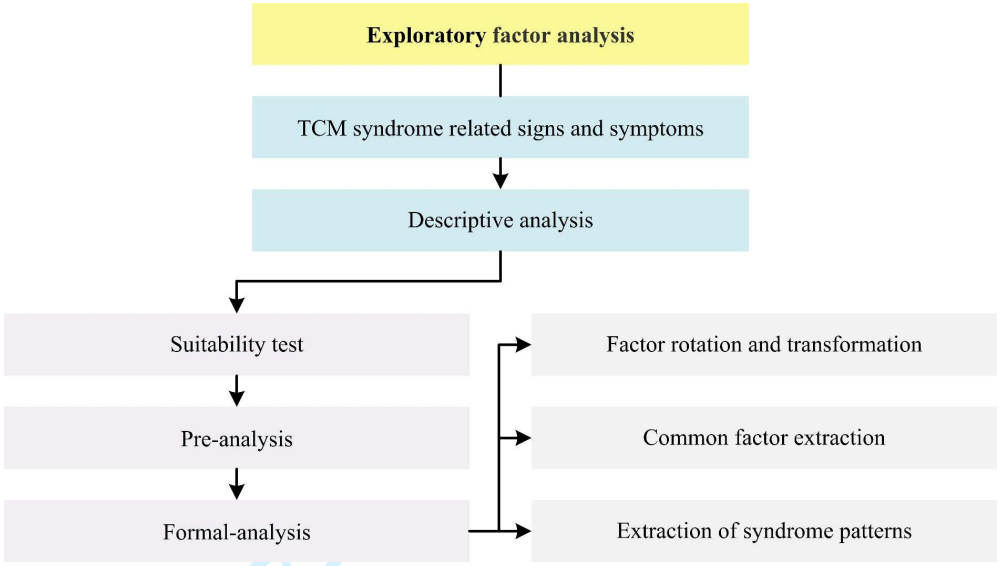


Fig.4 Flow diagram of association rule analysis in this study

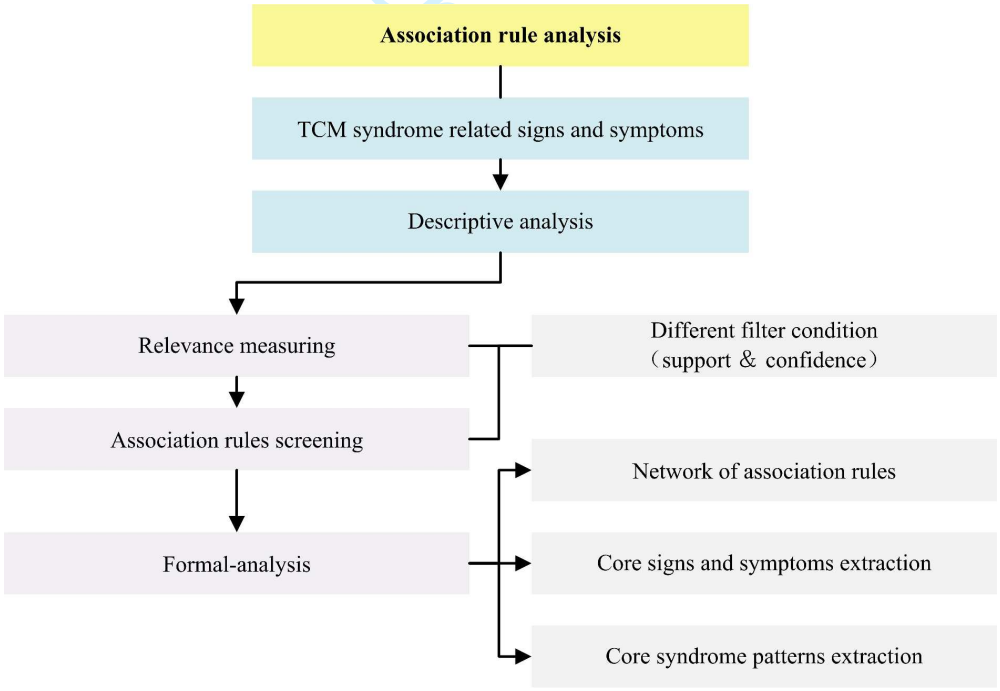
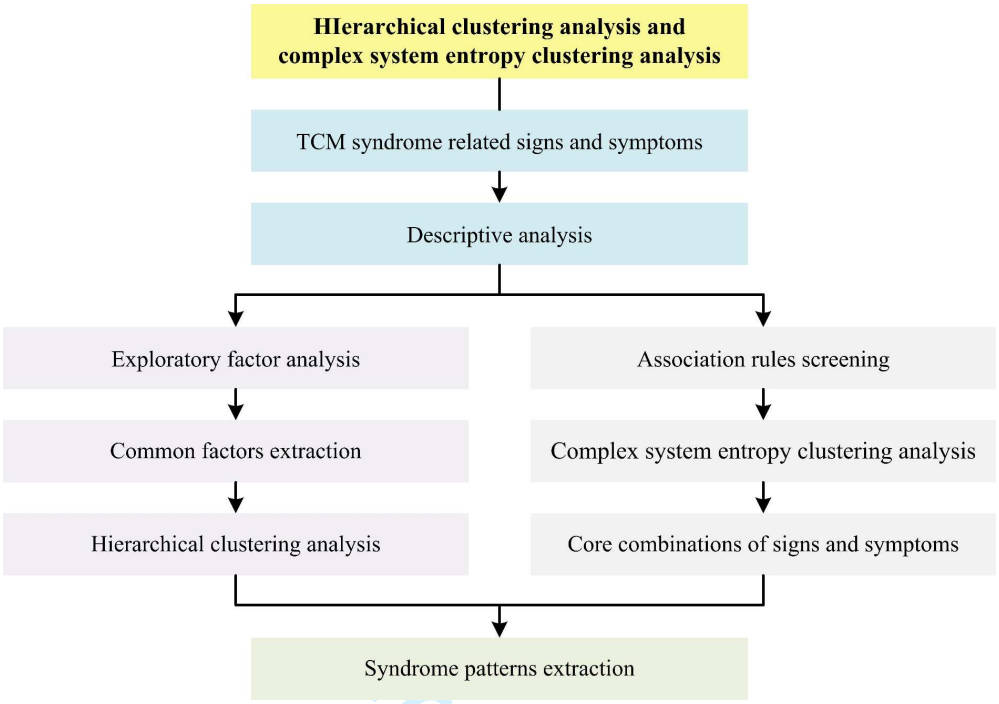


Fig.5 Flow diagram of hierarchical and complex system entropy clustering analysis in this study



Study registration and ethics

This study has been registered on ClinicalTrials.gov (No. NCT03314038). The protocol is designed in accordance with the Biological Ethics Review Method Involving Humans by the Ministry of Health of the People's Republic of China, and the Declaration of Helsinki. Ethical Review Board of Donzhimen Hospital Affiliated to Beijing University of Chinese Medicine has approved the protocol (No. ECPJ-BDY-2014-02). All the information will be used for only academic aims.

Dissemination

All the study outcomes will be disseminated on national conference reports, peer-reviewed TCM and modern medicine journals, or other academic publications. Findings from this study will not only provide much-needed understandings in this topic for both TCM clinical practitioners and researchers, but also greatly contribute to optimization of further prospective study design.

Strengths and limitations of this study

Strengths of our study should be addressed. Firstly, the current large cross-sectional study is the first research conducted to reveal TCM syndrome features in varied stages of CG malignant transformation. Although the current study design may limit generalisability of the findings, considering the study expenses of prospective study and the relatively low yearly malignant transformation rate of CG individuals, it is important for us to carry out this pilot study speculating potential TCM syndrome dynamic features from CG to GC. The study outcomes will provide targeted clues for the future design of long-term follow-up studies aiming to further explore novel risk prediction and TCM syndrome evolution models for CNAG or CAG malignant transformation combining both TCM and modern medicine indicators. Secondly, diagnosis of CNAG, CAG, GC, IM, GED and Hp infection in this study is confirmed based on latest gastroscope

reports, Hp test reports, and pathological reports at the time of recruitment. The accurate diagnostic information collected will greatly support conducting detailed subgroup analysis stratified by specific stages and pathological conditions (e.g., CAG with IM and/or GED, CAG with/without Hp infection, etc.). Thirdly, full-scale information collected in this study provides a good basis for in-depth exploration of clinical features. In addition, the parallelly applied biostatistical analysis strategies and unsupervised data mining strategies can reinforce complementary advantages and help us achieve a better understanding of TCM syndrome features in varied stages of the disease.

Our study also has several limitations. Firstly, this cross-sectional study merely involves data collection at a defined time (recruitment), dynamic characteristics of each invidual in the whole course of disease could thus not be tracked. Further large prospective studies with long follow-up time are highly demanded to expand findings of this study in a broaden range of settings, and thus contribute to building longitudinal data based novel risk assessment models of CG malignant transformation combining both TCM and modern medicine indicators. Secondly, as data are collected merely from TCM hospitals in Beijing, the current participants may not be representative enough to show the clinical features of patients from modern medicine hospitals, or medical centres in other regions of China. Selection bias may thus exist. Thirdly, some of the signs and symptoms related TCM indicators may be complicated enough for patients to fully understand and correctly report without prior TCM knowledge. A list of degree-definition of all the TCM indicators has been attached to each CRF for investigators and participants to refer to in recruitment, and each patient is interviewed by at least two TCM physicians to ensure quality control in this study. However, information bias may still inevitably exist to some extent, especially due to the complexity of individual self-reported TCM indicators.

Acknowledgments

This study was generously supported by grants from the National Natural Science Foundation of China (Xia Ding: No. 81630080, 91129714, 81270466, 81173424 and 81373796), National Undergraduates Innovating Experimentation Project of the Ministry of Education of China (Xia Ding: No. 081002609), Research Foundation of the Doctoral Program of Higher Education of China (Yin Zhang: No. 20120013110014), and Science Research Foundation of BUCM (Yin Zhang: No. 2014-JYBZZ-XS-134).

The authors assume full responsibility for study design, programming, data analyses and interpretation of the outcomes. All authors approved the final version of the manuscript. We are grateful to all the participants and staffs from Dongzhimen Hospital Affiliated to BUCM, Dongfang Hospital Affiliated to BUCM, Wangjing Hospital Affiliated to CACMS, and the 3rd Affiliated Hospital of BUCM for their remarkable contribution to this study.

Authors' contributions

Study concept: Xia Ding

Study and statistical analysis design design: Xia Ding, Yin Zhang, Yue Liu, Rui Song, and Zeqi Su

Drafting of the manuscript: Yin Zhang, Yue Liu, and Rui Song

Critical review and revisions of the manuscript: Xia Ding, Li Zhang, Yannan Li, Runhua Chen, Ning Shi, Xia Zhao, and Shiyu Du

Yin Zhang, Yue Liu and Rui Song contributed equally to this work.

Competing interests

The authors report no competing interests.

Abbreviations

CG: Chronic gastritis; CNAG: Chronic non-atrophic gastritis; CAG: Chronic atrophic gastritis; GC: Gastric carcinoma; TCM: Traditional Chinese medicine.

References

1. J. Y. Fang, W. Z. Liu, Z. S. Li, et al., "Consensus on chronic gastritis in China (version 2012)," *Chinese Journal of Gastroenterology*, vol. 18, no. 1, pp. 24-36, 2013. [Chinese]
2. M. Dinis-Ribeiro, M. Areia, A.C. de Vries, et al., "Management of precancerous conditions and lesions in the stomach (MAPS): guideline from the European Society of Gastrointestinal Endoscopy (ESGE), European Helicobacter Study Group (EHSG), European Society of Pathology (ESP), and the Sociedade Portuguesa de Endoscopia Digestiva (SPED)," *Endoscopy*, vol. 44, no. 1, pp. 74-94, 2012.
3. P. Correa, "A human model of gastric carcinogenesis," *Cancer Research*, vol. 48, no. 13, pp. 3554–3560, 1988.
4. P. Correa, "Human gastric carcinogenesis: a multistep and multifactorial process - First American Cancer Society Award Lecture on Cancer Epidemiology and Prevention," *Cancer Research*, vol. 52, no. 24, pp. 6735-6740, 1992.
5. L. A. Torre, F. Bray, R. L. Siegel, et al., "Global cancer statistics, 2012," *CA: A Cancer Journal for Clinicians*, vol. 65, no. 2, pp. 87-108, 2012.
6. Y. H. Park, N. Kim, "Review of atrophic gastritis and intestinal metaplasia as a premalignant lesion of gastric cancer," *Journal of Cancer Prevention*, vol. 20, no. 1, pp. 25-40, 2015.
7. P. Correa, M. B. Piazuelo, "The gastric precancerous cascade," *Journal of Digestive Diseases*, vol. 13, no. 1, pp. 2-9, 2012.
8. A. C. de Vries, N. C. van Grieken, C. W. Looman, et al., "Gastric cancer risk in patients with premalignant gastric lesions: a Nationwide Cohort Study in the Netherlands," *Gastroenterology*, vol. 134, no. 4, pp. 945-952, 2008.
9. W. J. Den Hollander, E.J. Kuipers, "Current pharmacotherapy options for gastritis," *Expert Opinion on Pharmacotherapy*, vol. 13, no. 18, pp. 2625-2636, 2012.
10. Y. K. Dai , Y. Z. Zhang, D.Y. Li, et al., "The efficacy of Jianpi Yiqi therapy for chronic atrophic gastritis: a systematic review and meta-analysis," *PLoS One*, vol. 12, no. 7: e0181906, 2017.
11. R. D. Cheng, B. Zhao, S. Q. Ma, et al., "Therapeutic effect of activating blood therapy in treating chronic atrophic gastritis: a meta-analysis," *China Journal of Traditional Chinese Medicine and Pharmacy*, vol. 32, no. 5, pp. 2050-2054, 2017. [Chinese]
12. Z. L. Guo, Z. N. Su, Z. F. Wang, et al., "Therapeutic effect of modified Sijunzi Tang in treating chronic atrophic gastritis: a meta-analysis," *Chinese Journal of Experimental Traditional Medical Formula*, vol. 21, no. 24, pp. 204-208, 2015. [Chinese]
13. J. X. Yang, J. An, J. S. Peng, et al., "Therapeutic effect of Banxia Xiexin decoction in treating chronic atrophic gastritis: a meta-analysis," *Journal of Beijing University of Chinese Medicine*, vol. 38, no. 1,

pp. 46-52, 2015. [Chinese]

14. W. L. Liao, G. Z. Chen, H. Hu, et al., "Therapeutic effect of Weifuchun in treating chronic atrophic gastritis: a meta-analysis," *Liaoning Journal of Traditional Chinese Medicine*, vol. 42, no. 1, pp. 11-15, 2015. [Chinese]
15. S.S. Zhang, Q.G. Li, H.P. Hang, et al., "Traditional Chinese medicine consensus on chronic superficial gastritis diagnosis and treatment (version 2009)", *Chinese Journal of Integrative Gastroenterology*, vol. 18, no. 3, pp. 207-209, 2010. [Chinese]
16. S.S. Zhang, Q.G. Li, X.D. Tang, et al., "Traditional Chinese medicine consensus on chronic atrophic gastritis diagnosis and treatment (version 2009)", *Chinese Journal of Integrative Gastroenterology*, vol. 18, no. 5, pp. 345-349, 2010. [Chinese]
17. W.D. Zhang, J.X. Li, Z.S. Chen, et al., "Integrative medicine consensus on chronic gastritis diagnosis and treatment in China (version 2012)", *Chinese Journal of Integrated Traditional and Western Medicine*, vol.32, no.6, pp.738-743, 2012. [Chinese]
18. National Health and Family Planning Commission of the People's Republic of China, "Diagnosis and treatment guideline for gastric carcinoma (trial version)," Available from URL: http://www.moh.gov.cn/ewebeditor/uploadfile/2013/05/20130515115755662.pdf
19. Z. Q. Su, P. C. Li, Q. Guo, et al., "Syndrome pattern transformation in chronic gastritis," *Journal of Beijing University of Traditional Chinese Medicine*, vol. 38, no. 11, pp. 762-771, 2015. [Chinese]
20. H. J. Gan, L. L. Liang, M. L. Han, S. R. Guo, X. J. Lin, "Relationship between syndrome distribution and chronic atrophic gastritis," *China Journal of Traditional Chinese Medicine and Pharmacy*, vol. 30, no. 7, pp. 2456-2459, 2015. [Chinese]
21. P. C. Li, X. Ding, Q. Guo, Z. Q. Su. "Literature review of Chinese medicine syndromes and syndrome element distribution features of chronic gastritis," *Liaoning Journal of Traditional Chinese Medicine*, vol. 41, no. 12, pp. 2532-2534, 2014. [Chinese]
22. W. J. Liu, Y. Xie, H. Cheng, et al. "Chinese consensus on Helicobacter pylori infection treatment (version 2012)," *Chinese Journal of Gastroenterology*, vol. 17, no. 10, pp. 3-21, 2014. [Chinese]
23. Q. Chen, H. Lu, "Kyoto global consensus report on Helicobacter pylori gastritis and its impact on Chinese clinical practice," *Journal of Digestive Diseases*, vol. 17, no. 6, pp. 353-356, 2016.
24. K. Sugano, J. Tack, E.J. Kuipers, et al., "Kyoto global consensus report on Helicobacter pylori gastritis," *Gut*, vol. 64, no. 9, pp. 1353-67, 2015.
25. P. Malfertheiner, F. Mégraud, C. A. O'Morain, et al., "Management of Helicobacter pylori infection - the Maastricht IV Florence Consensus Report," *Gut*, vol. 61, no. 5, pp. 646–64, 2012.
26. World Gastroenterology Organisation, "World Gastroenterology Organization Global Guideline: Helicobacter pylori in developing countries," *Journal of Clinical Gastroenterology*, vol. 45, no. 5, pp. 383-388, 2011.
27. H. Y. Hsiung, J. C. Fann, A. M. Yen, et al., "Stage-specific dietary factors associated with the Correa multistep and multifactorial process of human gastric carcinogenesis," *Nutrition and Cancer - An International Journal*, vol. 68, no. 4, pp. 598-610, 2016.

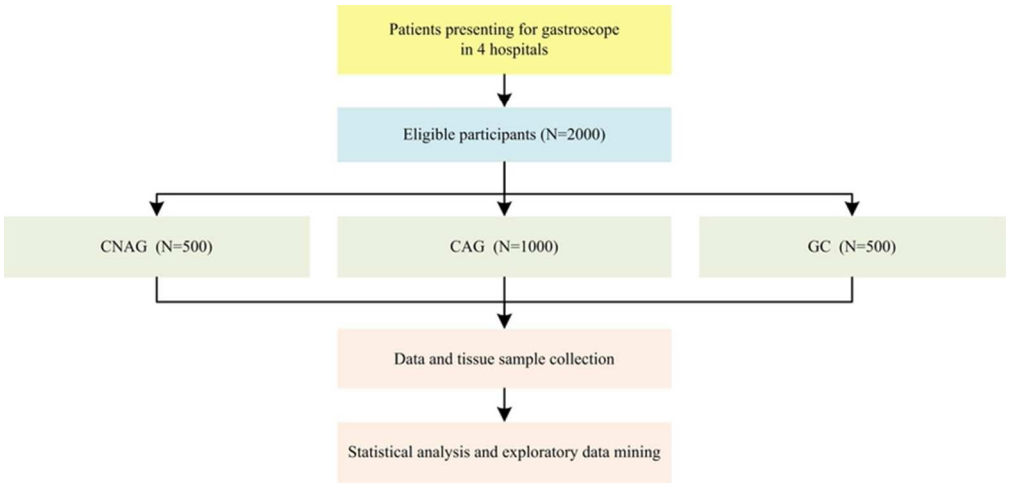


Fig.1 General flow diagram of this study

71x34mm (300 x 300 DPI)

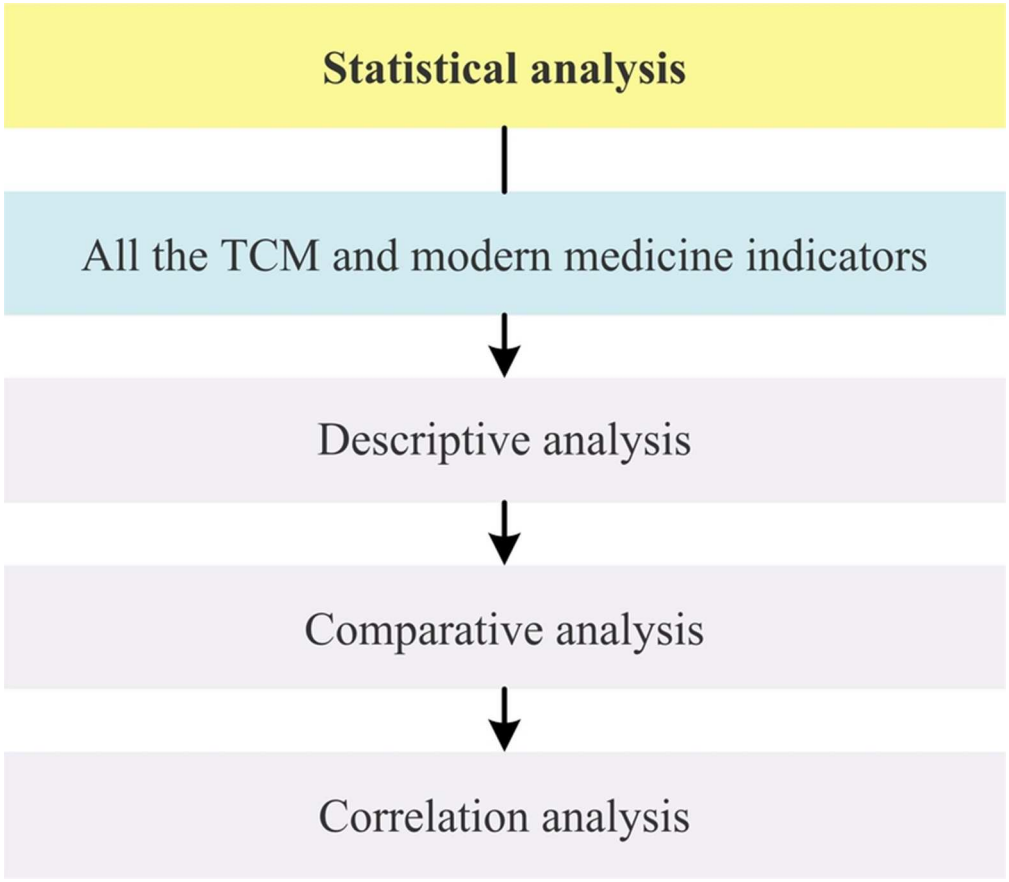


Fig.2 Flow diagram of statistical analysis in this study

70x61mm (300 x 300 DPI)

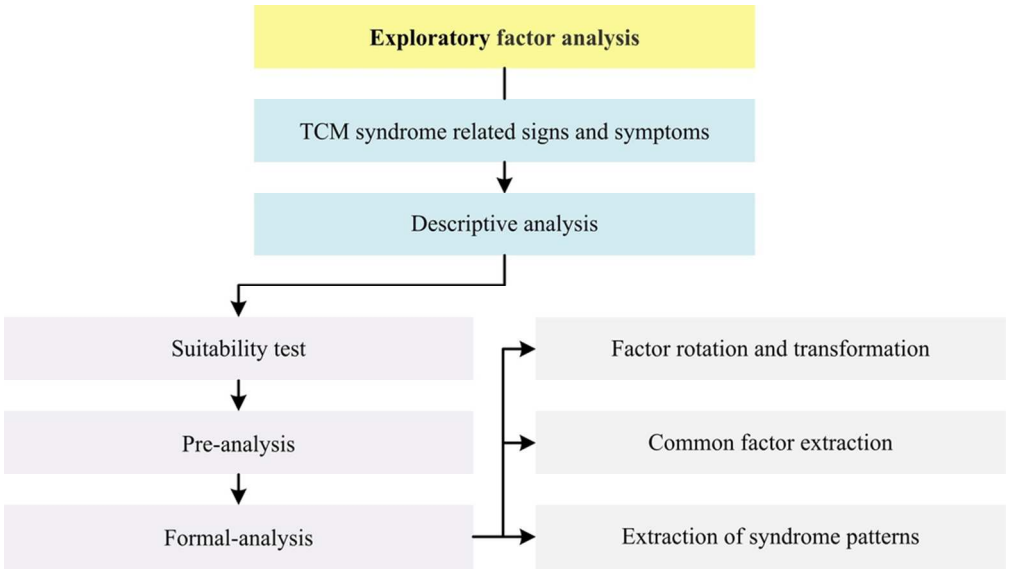


Fig.3 Flow diagram of exploratory factor analysis in this study

89x50mm (300 x 300 DPI)

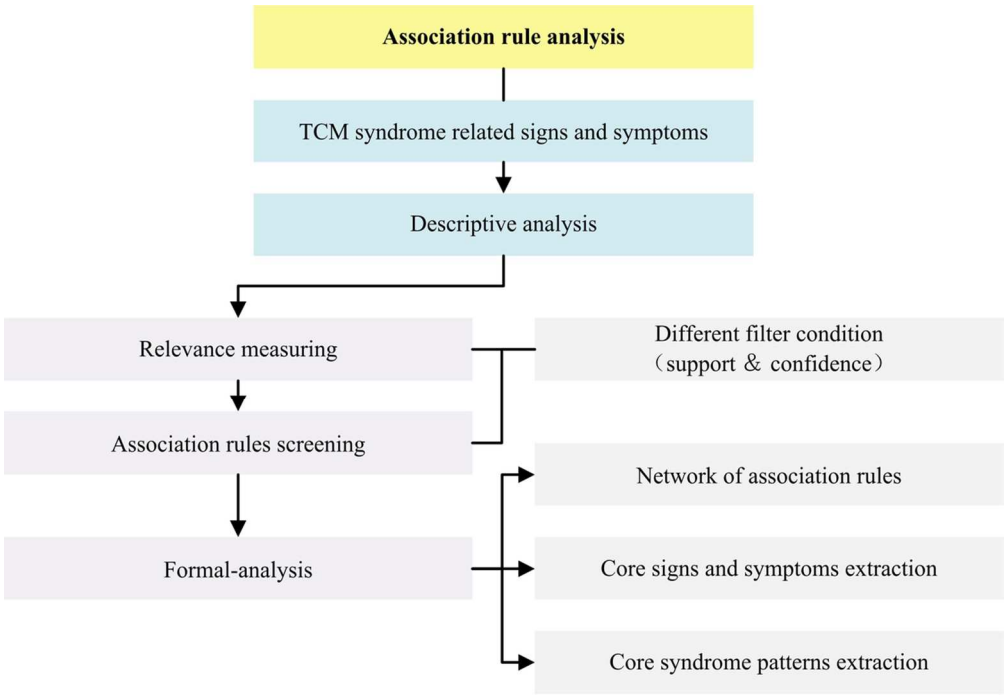


Fig.4 Flow diagram of association rule analysis in this study

110x75mm (300 x 300 DPI)

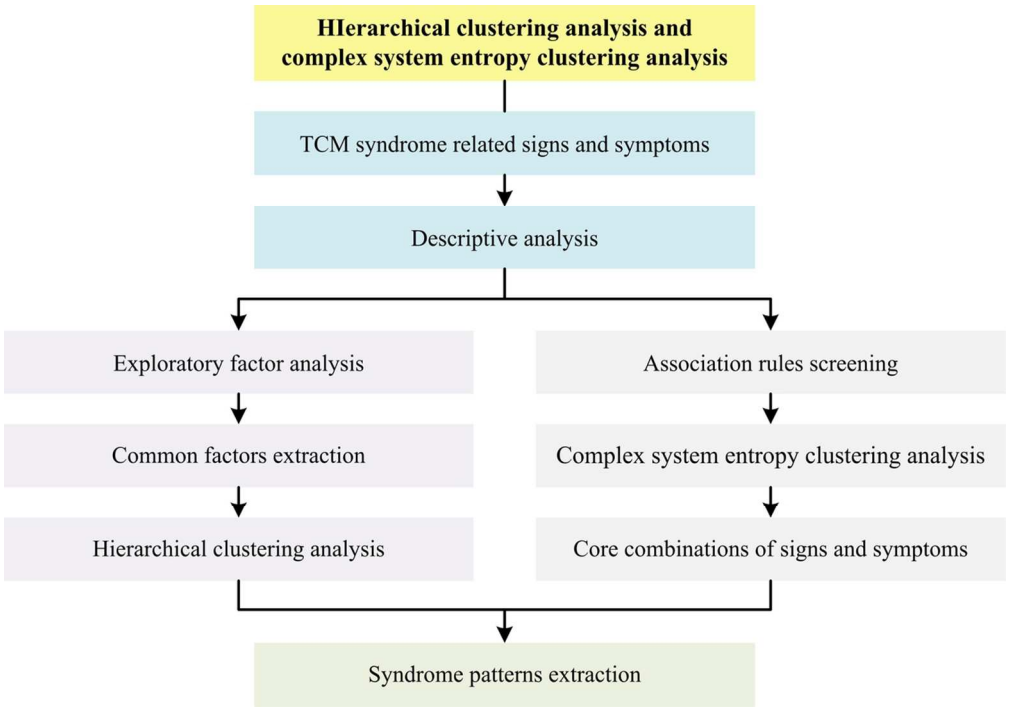


Fig.5 Flow diagram of hierarchical clustering analysis and complex system entropy clustering analysis in this study

112x78mm (300 x 300 DPI)

STROBE 2007 (v4) checklist of items to be included in reports of observational studies in epidemiology*

Checklist for cohort, case-control, and cross-sectional studies (combined)

Section/Topic	Item #	Recommendation	Reported on page #
Title and abstract	1	(*a*) Indicate the study's design with a commonly used term in the title or the abstract	1
		(*b*) Provide in the abstract an informative and balanced summary of what was done and what was found	1
Introduction			
Background/rationale	2	Explain the scientific background and rationale for the investigation being reported	3
Objectives	3	State specific objectives, including any pre-specified hypotheses	3,4
Methods			
Study design	4	Present key elements of study design early in the paper	4
Setting	5	Describe the setting, locations, and relevant dates, including periods of recruitment, exposure, follow-up, and data collection	4,5
Participants	6	(*a*) *Cohort study*—Give the eligibility criteria, and the sources and methods of selection of participants. Describe methods of follow-up *Case-control study*—Give the eligibility criteria, and the sources and methods of case ascertainment and control selection. Give the rationale for the choice of cases and controls *Cross-sectional study*—Give the eligibility criteria, and the sources and methods of selection of participants	5
		(*b*) *Cohort study*—For matched studies, give matching criteria and number of exposed and unexposed *Case-control study*—For matched studies, give matching criteria and the number of controls per case	Not applicable
Variables	7	Clearly define all outcomes, exposures, predictors, potential confounders, and effect modifiers. Give diagnostic criteria, if applicable	5,6
Data sources/ measurement	8*	For each variable of interest, give sources of data and details of methods of assessment (measurement). Describe comparability of assessment methods if there is more than one group	5,6
Bias	9	Describe any efforts to address potential sources of bias	4,5
Study size	10	Explain how the study size was arrived at	4
Quantitative variables	11	Explain how quantitative variables were handled in the analyses. If applicable, describe which groupings were chosen and why	7
Statistical methods	12	(*a*) Describe all statistical methods, including those used to control for confounding	7
		(*b*) Describe any methods used to examine subgroups and interactions	7
		(*c*) Explain how missing data were addressed	7
		(*d*) *Cohort study*—If applicable, explain how loss to follow-up was addressed *Case-control study*—If applicable, explain how matching of cases and controls was addressed	Not applicable

		Cross-sectional study—If applicable, describe analytical methods taking account of sampling strategy	
		(*e*) Describe any sensitivity analyses	Not applicable
Results			
Participants	13*	(a) Report numbers of individuals at each stage of study—eg numbers potentially eligible, examined for eligibility, confirmed eligible, included in the study, completing follow-up, and analysed	Not applicable (protocol)
		(b) Give reasons for non-participation at each stage	Not applicable (protocol)
		(c) Consider use of a flow diagram	4,6
Descriptive data	14*	(a) Give characteristics of study participants (eg demographic, clinical, social) and information on exposures and potential confounders	Not applicable (protocol)
		(b) Indicate number of participants with missing data for each variable of interest	Not applicable (protocol)
		(c) *Cohort study*—Summarise follow-up time (eg, average and total amount)	Not applicable (protocol)
Outcome data	15*	*Cohort study*—Report numbers of outcome events or summary measures over time	Not applicable (protocol)
		Case-control study—Report numbers in each exposure category, or summary measures of exposure	Not applicable (protocol)
		Cross-sectional study—Report numbers of outcome events or summary measures	Not applicable (protocol)
Main results	16	(*a*) Give unadjusted estimates and, if applicable, confounder-adjusted estimates and their precision (eg, 95% confidence interval). Make clear which confounders were adjusted for and why they were included	Not applicable (protocol)
		(*b*) Report category boundaries when continuous variables were categorized	Not applicable (protocol)
		(*c*) If relevant, consider translating estimates of relative risk into absolute risk for a meaningful time period	Not applicable (protocol)
Other analyses	17	Report other analyses done—eg analyses of subgroups and interactions, and sensitivity analyses	Not applicable (protocol)
Discussion			
Key results	18	Summarise key results with reference to study objectives	Not applicable (protocol)

Limitations	19	Discuss limitations of the study, taking into account sources of potential bias or imprecision. Discuss both direction and magnitude of any potential bias	10
Interpretation	20	Give a cautious overall interpretation of results considering objectives, limitations, multiplicity of analyses, results from similar studies, and other relevant evidence	9,10
Generalisability	21	Discuss the generalisability (external validity) of the study results	10
Other information			
Funding	22	Give the source of funding and the role of the funders for the present study and, if applicable, for the original study on which the present article is based	10

*Give information separately for cases and controls in case-control studies and, if applicable, for exposed and unexposed groups in cohort and cross-sectional studies.
Note: An Explanation and Elaboration article discusses each checklist item and gives methodological background and published examples of transparent reporting. The STROBE checklist is best used in conjunction with this article (freely available on the Web sites of PLoS Medicine at http://www.plosmedicine.org/, Annals of Internal Medicine at http://www.annals.org/, and Epidemiology at http://www.epidem.com/). Information on the STROBE Initiative is available at www.strobe-statement.org.

BMJ Open

Validating traditional Chinese syndrome features in varied stages of chronic gastritis malignant transformation: study protocol for a cross-sectional study

Journal:	*BMJ Open*
Manuscript ID	bmjopen-2017-020939.R1
Article Type:	Protocol
Date Submitted by the Author:	01-May-2018
Complete List of Authors:	Zhang, Yin; Bregham and Women's Hospital and Harvard Medical School, Medicine; Beijing University of Chinese Medicine, School of Chinese Medicine Liu, Yue; Beijing University of Chinese Medicine, School of Chinese Medicine Song, Rui; The Association of Chinese American Physicians; American Chinese Medical Exchange Society Zhang, Li; China-Japan Friendship Hospital, Department of Neurosurgery Su, Zeqi; Beijing University of Chinese Medicine, School of Chinese Medicine Li, Yannan; Affiliated Hospital of Shandong University of Chinese Medicine, Department of Gynecology Chen, Runhua; Dongfang Hospital Affiliated to Beijing University of Chinese Medicine, Department of Gastroenterology Shi, Ning; Beijing Electric Power Hospital, Department of Health Management Zhao, Xia; Beijing University of Chinese Medicine, School of Humanities Du, Shiyu; China-Japan Friendship Hospital, Department of Gastroenterology Ding, Xia; Beijing University of Chinese Medicine, School of Chinese Medicine
Primary Subject Heading:	Complementary medicine
Secondary Subject Heading:	Gastroenterology and hepatology
Keywords:	Chronic gastritis, Malignant transformation, Traditional Chinese medicine, Syndrome, Cross-sectional study

Validating traditional Chinese syndrome features in varied stages of chronic gastritis malignant transformation: study protocol for a cross-sectional study

Yin Zhang,[1,2,3,a] Yue Liu,[1,a] Rui Song,[4,5,a] Li Zhang,[6] Zeqi Su,[1] Yannan Li,[7] Runhua Chen,[8,1] Ning Shi,[9,1] Xia Zhao,[10] Shiyu Du,[11] and Xia Ding [1,*]

[1]School of Chinese Medicine, Beijing University of Chinese Medicine, Chaoyang District, Beijing 100029, China

[2]Channing Division of Network Medicine, Department of Medicine, Brigham and Women's Hospital and Harvard Medical School, Boston, MA 02115, USA

[3]Institute of Basic Research in Clinical Medicine, China Academy of Chinese Medical Sciences, Dongcheng District, Beijing 100700, China

[4]The Association of Chinese American Physicians, Flushing, NY 11354, USA

[5]American Chinese Medical Exchange Society, Burlington, MA 01803, USA

[6]Department of Neurosurgery, China-Japan Friendship Hospital, Chaoyang District, Beijing 100029, China

[7]Department of Gynecology, Affiliated Hospital of Shandong University of Chinese Medicine, Lixia District, Jinan 250011, China

[8]Department of Gastroenterology, Dongfang Hospital Affiliated to Beijing University of Chinese Medicine, Beijing 100078, China

[9]Department of Health Management, Beijing Electric Power Hospital, Fengtai District, Beijing 100073, China

[10]School of Humanities, Beijing University of Chinese Medicine, Chaoyang District, Beijing 100029, China

[11]Department of Gastroenterology, China-Japan Friendship Hospital, Chaoyang District, Beijing 100029, China

[a] These authors contributed equally to this work.

[*] Correspondence to: Dr. Xia Ding, E-mail: dingx@bucm.edu.cn

Word count: 2900 words (excluding title page, abstract, references, figures and tables)

Abstract

Introduction

The transition from chronic non-atrophic gastritis (CNAG) to chronic atrophic gastritis (CAG) and gastric carcinoma (GC) is regarded as a representative disease model of gastric mucosa malignant transformation leaded by uncontrolled inflammation. Traditional Chinese medicine (TCM) syndrome-targeted therapies have been applied in treating CG (chronic gastritis) malignant transformation in China with satisfying efficacy. This study aims to validate TCM syndrome features in each stage of CG malignant transformation. The findings may shed light on the TCM pathogenesis theory of CG malignant transformation, and thus provides potential translation for optimization of syndrome-targeted treatment strategies of CNAG, CAG, and GC respectively.

Methods and analysis

The present study is a cross-sectional study conducted in China. 2000 eligible patients, including 500 CNAG cases, 1000 CAG cases, and 500 GC cases, will be recruited from 4 TCM hospitals. Descriptive analysis, comparative analysis, and correlation analysis of all the measurement data will be performed by biostatisticians. Unsupervised data mining analyses, including exploratory factor analysis, association rule analysis, hierarchical clustering analysis, and complex system entropy clustering analysis, etc., will also be performed by data scientists respectively for indepth analyses of TCM syndrome-related indicators.

Ethics and dissemination

The protocol has been approved by Ethical Review Board of Donzhimen Hospital Affiliated to Beijing University of Chinese Medicine (No. ECPJ-BDY-2014-02). All the study outcomes will be disseminated through national conference reports and meantime published on peer-reviewed journals.

Study registration

This study has been registered on ClinicalTrials.gov (No. NCT03314038).

Keywords:

Chronic gastritis; Malignant transformation; Traditional Chinese medicine; Syndrome; Cross-sectional study

Strengths and limitations of this study

►To the authors' knowledge, this is among the first reported large population based studies on validating TCM syndrome features in varied stages of CG malignant transformation.

►A large sample size of 2000 participants can be obtained. Full-scale TCM and modern medicine indicators can be accurately collected.

►Both statistical analysis and unsupervised data mining strategies will be applied by research personnel parallelly to reinforce complementary advantages.

►This is not a prospective study, and thus longitudinal data couldn't be collected by research personnel to speculate causal relationship.

►Selection bias and information bias may exist.

Introduction

Chronic gastritis (CG), persistent inflammation of the gastric mucosa with no specific clinical manifestations, has been regarded as the most common gastrointestinal disease with an incidence of 70% among the adult population in China [1]. Chronicity and recurrence of chronic non-atrophic gastritis (CNAG) may lead to chronic atrophic gastritis (CAG), which is characterized by gastric mucosal atrophy and usually accompany with metaplasia or dysplasia [2]. As the most important intermediate step of CG malignant transformation, CAG is acknowledged as premalignant lesion of gastric cancer (GC). The sequential stages of inflammation, metaplasia, dysplasia and carcinoma have been proved a well-established tumorigenesis model of intestinal type of GC leaded by uncontrolled inflammation [2-4]. Statistics have shown that there are 951,600 annual new cases of GC worldwide [5]. China has been reported as one of the countries with largest population of both CG and GC sufferers [1]. Considering the severe burden of above-mentioned diseases, active intervention on CG malignant transformation process is highly needed [6-8].

Traditional Chinese medicine (TCM), as one of the most important and time-honored alternative medicine approaches worldwide, has developed unique theories of etiology and diagnosis since ancient time. According to TCM theory, syndrome is defined as a categorised pattern of symptoms and signs in a patient at a specific stage during the course of specific diseases, and is considered the most important unit for evaluating pathogenesis. TCM practitioners have been applying the central principles of syndrome pattern differentiation in clinical practice to identify the physical condition of patients and the pathologensis of diseases with a history of more than 2500 years. Nowadays, TCM physicians keep on using the prestigious and traditional "four diagnostic methods" of looking, listening/smelling, asking and palpating to collect the information of symptoms and signs on patients comprehensively. Full-scale assessment of aforementioned clinical information can then be conducted by physicians based on TCM theory and clinical experience to reveal the etiology (syndrome) of diseases, and thus guiding individualized herbal prescriptions. Though huge leaps have been seen in past decades in elaborating the pathogenesis of CAG, however, modern medicines in treating CNAG and CAG remain unsatisfied to some extent [9]. Owing to the heavy burden of recurrence, chronicity, and malignant transformation of CG, a considerable proportion of sufferers in China have put their concentrations on TCM therapies.

Trials have indicated the benefits of TCM syndrome-targeted interventions on CNAG and CAG [10-14]. The Society of Gastroenterology, China Association of Chinese Medicine has propagated the latest "*TCM Consensus on Chronic Superficial Gastritis Diagnosis and Treatment*" and "*TCM Consensus on Chronic Atrophic Gastritis Diagnosis and Treatment*" respectively in 2009 [15, 16]. The Society of Gastroenterology, China Society of Integrated Traditional Chinese and Western Medicine also issued "*Integrative Medicine Consensus on Chronic Gastritis Diagnosis and Treatment in China*" in 2012 [17]. The aforementioned consensuses emphasized the importance of applying TCM syndrome-targeted therapies in controlling gastric mucosal inflammation and atrophy, and inhibiting metaplasia and dysplasia in accordance with China's local conditions. "*Consensus on chronic gastritis in China*" promulgated by the Society of Gastroenterology, Chinese Medical Association and "*Diagnosis and Treatment Guideline for Gastric Carcinoma*" issued by the National Health and Family Planning Commission of the People's Republic of China also recommended the application of TCM syndrome differentiation based therapies in addition to conventional treatment respectively [1, 18].

Both consensus and literature reviews have also indicated the potential links between syndrome

patterns and CG malignant transformation [15-17, 19-21]. We believe that the diversity and dynamics of TCM syndromes may play a role all the way through this course. However, no large population based study has been reported to support in-depth analysis of prevalence and severity of varied TCM syndromes in different stages of CG malignant transformation. Taking note of the limitations of previous studies, we highlight validating detailed characteristics of TCM syndromes in CNAG, CAG, and GC population respectively in the present study. The new findings may shed light on the TCM pathogenesis theory of CG malignant transformation, and thus provide potential translation for optimization of syndrome-targeted treatment strategies by TCM herbal prescriptions in varied stages of this process.

Study aims

The primary motivation for the present study is to identify TCM syndrome features in varied stages of CG malignant transformation respectively, including CNAG, CAG, and GC. In addition, this study can also be regarded as a pilot study speculating potential dynamic features of TCM syndrome in the whole course of CG. This may provide important and targeted clues for the design of future observational long-term follow-up studies aiming to further explore novel risk prediction and syndrome evolution models combining both TCM and modern medicine indicators for CG or CAG malignant transformation.

Methods

Design

This is a cross-sectional study undertaken from 2014 to 2018, comprising 2000 eligible participants in 4 hospitals in China. The general flow diagram of this study is outlined in Fig. 1.

Fig.1 General flow diagram of this study

Recruitment

This study is leaded by Beijing University of Chinese Medicine (BUCM). The 4 medical centres participating in the present study include Dongzhimen Hospital Affiliated to BUCM, Dongfang Hospital Affiliated to BUCM, the 3rd Affiliated Hospital of BUCM, and Wangjing Hospital Affiliated to China Academy of Chinese Medical Sciences (CACMS). To make a definitive diagnosis for each participant, an upper gastrointestinal endoscopy based screening process is conducted by study personnel during recruitment. All the eligible patients presenting to the the aforementioned hospitals are then identified and recruited continuously until the required sample size for each category (CNAG, CAG, and GC) is achieved respectively. This study has opened to recruitment in 2014. Recruitment is ongoing and is expected to conclude by the end of 2018. The sample size required is calculated using the following formula: $N=(Z^2 p(1-p))/d^2$. Where N: sample size; P: expected proportion of certain syndrome, if 50%, P=0.5; d: precision, if 5%, d=0.05, which is conventional; Z statistic (Z): the level of confidence of 95%, conventionally, Z value is 1.96 for 95% CI. According to this formaula, the sample size would be 385 cases for each category (CNAG, CAG, and GC). However, taking into full consideration of items of CRFs, experts' opinions, study expenses, and operability, the anticipated sample size is finally decided to be 2000 in total, including 500 CNAG cases, 1000 CAG cases, and 500 GC cases.

Diagnostic criteria

Diagnostic criteria of CNAG, CAG, GMA, IM, and GED in the present study refers to "Consensus on

chronic gastritis in China" promulgated by the Society of Gastroenterology, Chinese Medical Association [1]. Diagnostic criteria of GC refers to "*Diagnosis and Treatment Guideline for Gastric Carcinoma (trail version)*" issued by the National Health and Family Planning Commission of the People's Republic of China [18]. Diagnostic criteria of Helicobacter pylori (Hp) infection refers to "*Chinese Consensus on Helicobacter pylori Infection Treatment*" issued by the Chinese Society of Gastroenterology, Chinese Medical Association [22]. Criteria for TCM syndrome pattern differentiation diagnosis in this study refers to "*TCM Consensus on Chronic Superficial Gastritis Diagnosis and Treatment*" [15] and "*TCM Consensus on Chronic Atrophic Gastritis Diagnosis and Treatment*" [16] issued by the Society of Gastroenterology, China Association of Chinese Medicine.

Inclusion criteria

The inclusion criteria are as follows: (1) Meeting diagnostic criteria of CNAG, CAG, or GC; (2) Willing to cooperate with investigators for data and tissue sample collection; (3) Willing to sign informed consent.

Exclusion criteria

The exclusion criteria are as follows: (1) History of previous stomach surgeries; (2) Unable to participate in data and sample collection for any reason.

Data and samples collection

Strict training in standard operating procedure (SOP) of clinical data and samples collection is conducted by principal investigator for research personnel before study initiation. Prior to recruitment, participants are required to sign informed consents. In this process, clinial investigators take the responsibility for explaining the objectives, general procedures, data collection methods, risks and benefits, authorization, data privacy policy, and other necessary details to every patient. At the time of the data collection the investigators are blinded for the groups the patients may be included. Research personnel will then assess all the indicators and tissue samples needed at once in recruitment. Each participant is interviewed by at least two physicians to ensure quality control. A list of degree-definition of all the TCM indicators is attached to each CRF for investigators and participants to refer to.

Outcome measures and overview of CRFs

Primary measures of this study include the prevalence of TCM syndrome patterns in varied stages of CG malignant transformation (CNAG, CAG, and GC). Secondary outcome measures include prevalence and severity of all the presenting signs and symptoms collected by using TCM four diagnostic methods (looking, listening/smelling, asking, and palpating). The aforementioned clinical features are identified in detail base on TCM syndrome pattern differentiation scales.

Accurate modern medicine diagnostic information in this study is achieved based on latest gastroscope reports, and pathological reports. Hp infects more than half of the populations worldwide and causes long-term progressive damage to the gastric mucosa in all infected individuals, is acknowledged as one of the most important risk factors for CG and GC [23-26]. Stage-specific dietary intake, ingestion of ascorbic acid and nitrate, and other factors have also been proved to associate with the multistep and multifactorial process of gastric carcinogenesis [1-4, 27]. Considering the importance of above-mentioned indicators, full-scale information regarding to Hp test reports, dietary intake, life behaviors, and medication, etc., are

collected in addition to TCM indicators. Other necessary information collected in our study covers demographic information, disease history, complications, and laboratory indicators. In addition, tissue samples collection is also included. Subgroup analysis can thus be carried out in a number of specific populations (e.g. CAG with metaplasia and/or dysplasia, CAG with/without Hp infection, etc.), which is of significant clinical value.

According to study design and case report forms (CRFs), detailed content of data and tissue samples collection are illustrated in Table. 1.

Table.1 Content of data and tissue samples collection

Content	Details
Demographic Information	Name, date of birth, gender, age, nationality, height, weight, level of education, etc.
Diagnostic Information	Latest gastroscope reports, Hp test reports, and pathological reports
TCM Syndrome Related Information	Features of presenting signs and symptoms collected by using TCM four diagnostic methods of looking, listening/smelling, asking, and palpating
	TCM syndrome pattern differentiation scale scores
	Results of TCM syndrome identification
Disease History	Past history and family history of CG, GC, and Hp infection, etc.
Complications	Upper gastrointestinal hemorrhage, upper gastrointestinal ulcer, pyloric obstruction, anemia, gastric carcinoma metastasis, etc.
Dietary Intake and Life Behaviors	Dietary structure, working state, psychological state, alcohol intake history, history of smoking, etc.
Medication	Current drug use on CG and GC treatment (current use of proton pump inhibitors, etc. at the time of upper endoscopy), and nonsteroidal anti-inflammatory drugs (NSAIDs)
Laboratory Indicators	Routine blood test, routine stool test, tumor markers, etc.
Tissue Samples	Tissue samples of diseased gastric mucosa

Patient and public involvement

Before the formal recruitment started, a pre-test of CRF was conducted by research personnel in patients at the Dongzhimen Hospital Affiliated to BUCM. In this process, principal investigator and research personnel collated the feedback from the patients and using it to improve the final design of the CRF. Study outcomes will be disseminated on conference reports and academic publications. Main findings will also be disseminated to study participants via email and the BUCM website.

Data management

SOP of data collection, entry, editing, locking and retrieval is set up by data management centre of this study in Beijing University of Chinese Medicine. Confidentiality, authenticity, and integrity of all the

clinical information gathered in the present study is highlighted and maintained at all levels of data management. CRFs are organized by unique identification numbers and thus no patients' personal identifiers will be available in any records. Before data-entry, double-checking of range and logic of every variable value is performed by primary researchers. Data is then double-entered into a dedicated database by data administrators to confirm there is no input error.

Data analysis and outcome reporting

In order to achieve the primary goal of validating TCM syndrome features in varied stages of CG malignant transformation, varied data analysis strategies on multidimensional indicators will be performed parallelly from both biostatistics and data mining perspectives.

Descriptive statistics will be carried out by biostatisticians for all the measurement data, including demographic and clinical characteristics in the whole study population and each stratum. Comparative analysis of qualitative TCM and modern medicine indicators in different groups will be performed using chi-squared tests. Comparative analysis of quantitative scores of TCM syndrome pattern differentiation scales will be conducted using rank-sum test. A probability of $P < 0.05$ will be considered statistically significant for these analyses. Correlation analysis of pathological grades and stages of CG malignant transformation and TCM indicators will be performed using logistic regression model. The significance level for introducing and removing variables will be 0.05 and 0.10.

In addition, unsupervised data mining analyses, including exploratory factor analysis, association rule analysis, hierarchical clustering analysis, and complex system entropy clustering analysis, etc., will be performed by data scientists respectively for indepth exploring of TCM syndrome distribution and evolution features in the whole course of CG malignant transformation.

Statistical analyses in this study will be carried out using SAS software *(version 9.3, SAS Institute Inc., Cary, NC, U.S.A)* , and R software *(version 3.3.1, The R Project for Statistical Computing., Auckland, N.Z.)*. Data mining analyses in this study will be processed by SPSS Clementine software *(version 12.0, SPSS Inc., Chicago, IL, U.S.A)*, and Matlab Software *(version 2016a, Mathworks Inc., Natick, MA, U.S.A)*. Plotting in this study will be performed using Office Excel software *(version 2007, Microsoft Inc., Redmond, WA, U.S.A)* and Office Visio software *(version 2007, Microsoft Inc., Redmond, WA, U.S.A)*.

General flow diagram of biostatistical ananlysis process is outlined in Fig. 2. Detailed flow diagrams of aforementioned exploratory data mining strategies to be conducted in this study are illustrated respectively in Fig. 3-5.

Fig.2 Flow diagram of statistical analysis strategy in this study

Fig.3 Flow diagram of exploratory factor analysis in this study

Fig.4 Flow diagram of association rule analysis in this study

Fig.5 Flow diagram of hierarchical and complex system entropy clustering analysis in this study

Study registration and ethics

This study has been registered on ClinicalTrials.gov (No. NCT03314038). The protocol is designed in accordance with the Biological Ethics Review Method Involving Humans by the Ministry of Health of the People's Republic of China, and the Declaration of Helsinki. Ethical Review Board of Donzhimen Hospital Affiliated to Beijing University of Chinese Medicine has approved the protocol (No. ECPJ-BDY-2014-02). All the information will be used for only academic aims.

Dissemination

All the study outcomes will be disseminated on national conference reports, peer-reviewed TCM and modern medicine journals, or other academic publications. Findings from this study will not only provide much-needed understandings in this topic for both TCM clinical practitioners and researchers, but also greatly contribute to optimization of further prospective study design.

Strengths and limitations of this study

Strengths of our study should be addressed. Firstly, the current large cross-sectional study is the first research conducted to reveal TCM syndrome features in varied stages of CG malignant transformation. Although the current study design may limit generalisability of the findings, considering the study expenses of prospective study and the relatively low yearly malignant transformation rate of CG individuals, it is important for us to carry out this pilot study speculating potential TCM syndrome dynamic features from CG to GC. The study outcomes will provide targeted clues for the future design of long-term follow-up studies aiming to further explore novel risk prediction and TCM syndrome evolution models for CNAG or CAG malignant transformation combining both TCM and modern medicine indicators. Secondly, diagnosis of CNAG, CAG, GC, IM, GED and Hp infection in this study is confirmed based on latest gastroscope reports, Hp test reports, and pathological reports at the time of recruitment. The accurate diagnostic information collected will greatly support conducting detailed subgroup analysis stratified by specific stages and pathological conditions (e.g., CAG with IM and/or GED, CAG with/without Hp infection, etc.). Thirdly, full-scale information collected in this study provides a good basis for in-depth exploration of clinical features. In addition, the parallelly applied biostatistical analysis strategies and unsupervised data mining strategies can reinforce complementary advantages and help us achieve a better understanding of TCM syndrome features in varied stages of the disease.

Our study also has several limitations. Firstly, this cross-sectional study merely involves data collection at a defined time (recruitment), dynamic characteristics of each invidual in the whole course of disease could thus not be tracked. Further large prospective studies with long follow-up time are highly demanded to expand findings of this study in a broaden range of settings, and thus contribute to building longitudinal data based novel risk assessment models of CG malignant transformation combining both TCM and modern medicine indicators. Secondly, as data are collected merely from TCM hospitals in Beijing, the current participants may not be representative enough to show the clinical features of patients from modern medicine hospitals, or medical centres in other regions of China. Selection bias may thus exist. Thirdly, some of the signs and symptoms related TCM indicators may be complicated enough for patients to fully understand and correctly report without prior TCM knowledge. A list of degree-definition of all the TCM indicators has been attached to each CRF for investigators and participants to refer to in recruitment, and each patient is interviewed by at least two TCM physicians to ensure quality control in this study. However, information bias may still inevitably exist to some extent, especially due to the complexity of

individual self-reported TCM indicators.

Acknowledgments

This study was generously supported by grants from the National Natural Science Foundation of China (Xia Ding: No. 81630080, 91129714, 81270466, 81173424 and 81373796), National Undergraduates Innovating Experimentation Project of the Ministry of Education of China (Xia Ding: No. 081002609), Research Foundation of the Doctoral Program of Higher Education of China (Yin Zhang: No. 20120013110014), and Science Research Foundation of BUCM (Yin Zhang: No. 2014-JYBZZ-XS-134).

The authors assume full responsibility for study design, programming, data analyses and interpretation of the outcomes. All authors approved the final version of the manuscript. We are grateful to all the participants and staffs from Dongzhimen Hospital Affiliated to BUCM, Dongfang Hospital Affiliated to BUCM, Wangjing Hospital Affiliated to CACMS, and the 3rd Affiliated Hospital of BUCM for their remarkable contribution to this study.

Authors' contributions

Study concept: Xia Ding

Study and statistical analysis design design: Xia Ding, Yin Zhang, Yue Liu, Rui Song, and Zeqi Su

Drafting of the manuscript: Yin Zhang, Yue Liu, and Rui Song

Critical review and revisions of the manuscript: Xia Ding, Li Zhang, Yannan Li, Runhua Chen, Ning Shi, Xia Zhao, and Shiyu Du

Yin Zhang, Yue Liu and Rui Song contributed equally to this work.

Competing interests

The authors report no competing interests.

Abbreviations

CG: Chronic gastritis; CNAG: Chronic non-atrophic gastritis; CAG: Chronic atrophic gastritis; GC: Gastric carcinoma; TCM: Traditional Chinese medicine.

References

1. Fang YJ, Liu WZ, Li ZS, *et al*. Consensus on chronic gastritis in China (version 2012). *Chin J Gastroenterol* 2013;18:24-36. [Chinese]
2. Dinis-Ribeiro M, Areia M, De Vries AC, *et al*. Management of precancerous conditions and lesions in the stomach (MAPS): guideline from the European Society of Gastrointestinal Endoscopy (ESGE), European Helicobacter Study Group (EHSG), European Society of Pathology (ESP), and the Sociedade Portuguesa de Endoscopia Digestiva (SPED). *Endoscopy* 2012;44:74-94.
3. Correa P. A human model of gastric carcinogenesis. *Cancer Res* 1988;48:3554–60,.
4. Correa P. Human gastric carcinogenesis: a multistep and multifactorial process - First American Cancer Society Award Lecture on Cancer Epidemiology and Prevention. *Cancer Res* 1992;52:6735-40.
5. Torre LA, Bray F, Siegel RL, *et al*. Global cancer statistics, 2012. *CA-Cancer J Clin* 2012;65:87-108.
6. Park YH, Kim N. Review of atrophic gastritis and intestinal metaplasia as a premalignant lesion of

gastric cancer. *J Cancer Prev* 2015;20:25-40.

7. Correa P, Piazuelo MB. The gastric precancerous cascade. *J Digest Dis* 2012;13:2-9.
8. De Vries AC, Van Grieken NC, Looman CW, *et al*. Gastric cancer risk in patients with premalignant gastric lesions: a Nationwide Cohort Study in the Netherlands. *Gastroenterology* 2008;134:945-52,.
9. Den Hollander WJ, Kuipers EJ. Current pharmacotherapy options for gastritis. *Expert Opin Pharmaco* 2012;13:2625-36.
10. Dai YK, Zhang YZ, Li DY, *et al*. The efficacy of Jianpi Yiqi therapy for chronic atrophic gastritis: a systematic review and meta-analysis. *PLoS One* 2017;12:e0181906.
11. Cheng RD, Zhao B, Ma SQ, *et al*. Therapeutic effect of activating blood therapy in treating chronic atrophic gastritis: a meta-analysis. *Chin J Tradit Chin Med Pharm* 2017;32:2050-4. [Chinese]
12. Guo ZL, Su ZN, Wang ZF, *et al*. Therapeutic effect of modified Sijunzi Tang in treating chronic atrophic gastritis: a meta-analysis," *Chin J Exp Tradit Med* 2015;21:204-8. [Chinese]
13. Yang JX, An J, Peng JS, *et al*. Therapeutic effect of Banxia Xiexin decoction in treating chronic atrophic gastritis: a meta-analysis. *J Beijing Univ Chin Med* 2015;38:46-52. [Chinese]
14. Liao WL, Chen GZ, Hu H, *et al*. Therapeutic effect of Weifuchun in treating chronic atrophic gastritis: a meta-analysis. *Liaoning J Tradit Chin Med* 2015;42:11-15. [Chinese]
15. Zhang SS, Li QG, Hang HP, *et al*. Traditional Chinese medicine consensus on chronic superficial gastritis diagnosis and treatment (version 2009). *Chin J Integr Gastroenterol* 2010; 18:207-9. [Chinese]
16. Zhang SS, Li QG, Tang XD, *et al*. Traditional Chinese medicine consensus on chronic atrophic gastritis diagnosis and treatment (version 2009). *Chin J Integr Gastroenterol* 2010;18:345-9. [Chinese]
17. Zhang WD, Li JX, Chen ZS, *et al*. Integrative medicine consensus on chronic gastritis diagnosis and treatment in China (version 2012). *Chin J Integr Tradit West Med* 2012;32:738-43. [Chinese]
18. National Health and Family Planning Commission of the People's Republic of China. "Diagnosis and treatment guideline for gastric carcinoma (trial version). Available from URL: http://www.moh.gov.cn/ewebeditor/uploadfile/2013/05/20130515115755662.pdf
19. Su ZQ, Li PC, Guo Q, *et al*. Syndrome pattern transformation in chronic gastritis. *J Beijing Univ Tradit Chin Med* 2015;38:762-71. [Chinese]
20. Gan HJ, Liang LL, Han ML, *et al*. Relationship between syndrome distribution and chronic atrophic gastritis. *Chin J Tradit Chin Med Pharm* 2015;30:2456-9. [Chinese]
21. Li PC, Ding X, Guo Q, *et al*. Literature review of Chinese medicine syndromes and syndrome element distribution features of chronic gastritis. *Liaoning J Tradit Chin Med* 2014;41:2532-34. [Chinese]
22. Liu WJ, Xie Y, Cheng H, *et al*. Chinese consensus on Helicobacter pylori infection treatment (version 2012). *Chin J Gastroenterol* 2014;17:3-21. [Chinese]
23. Chen Q, Lu H. Kyoto global consensus report on Helicobacter pylori gastritis and its impact on Chinese clinical practice. *J Digest Dis* 2016;17:353-6.
24. Sugano K, Tack J, Kuipers EJ, *et al*. Kyoto global consensus report on Helicobacter pylori gastritis. *Gut* 2015;64:1353-67.

25. Malfertheiner P, Mégraud F, O'Morain CA, *et al*. Management of Helicobacter pylori infection - the Maastricht IV Florence Consensus Report. *Gut* 2012;61:646–64.

26. World Gastroenterology Organisation. World Gastroenterology Organization Global Guideline: Helicobacter pylori in developing countries. *J Clin Gastroenterol* 2011;45:383-8.

27. Hsiung HY, Fann JC, Yen AM, *et al*. Stage-specific dietary factors associated with the Correa multistep and multifactorial process of human gastric carcinogenesis. *Nutr Cancer* 2016;68:598-610.

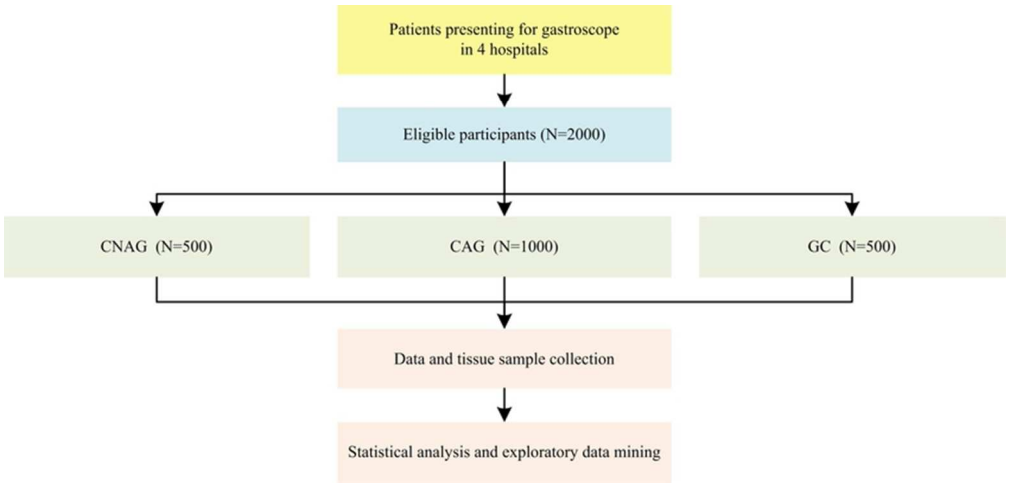


Fig.1 General flow diagram of this study

71x34mm (300 x 300 DPI)

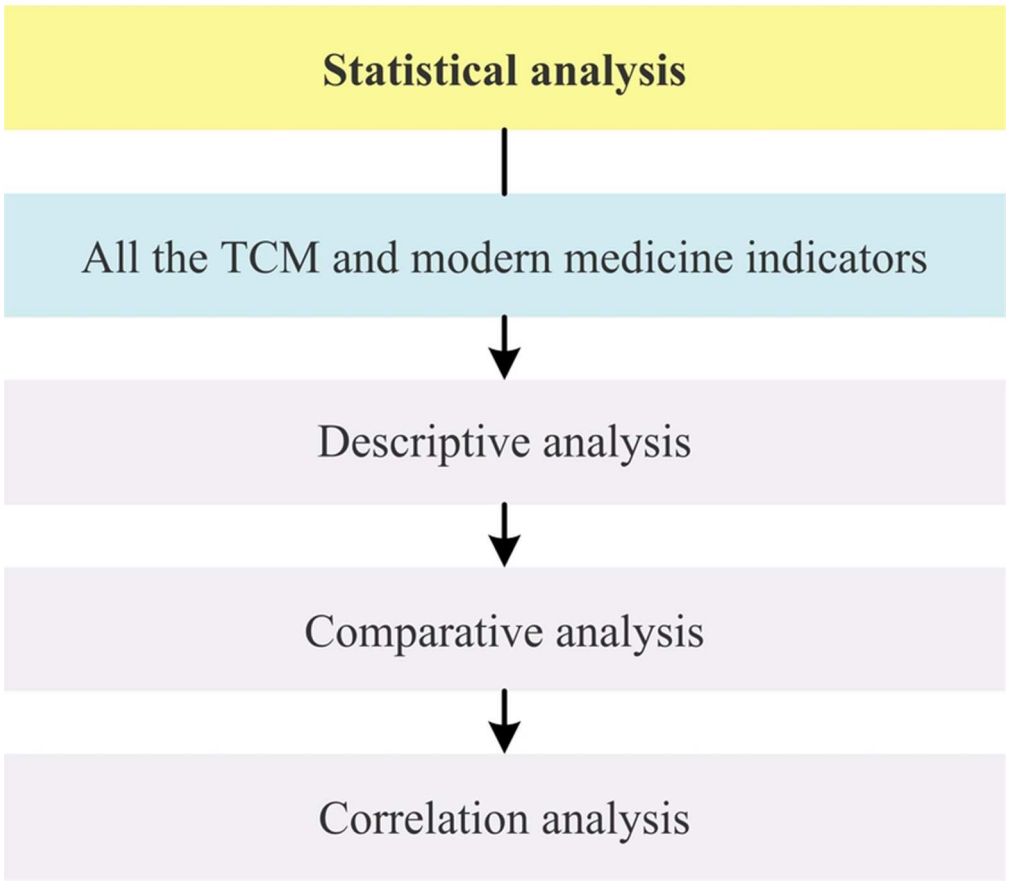


Fig.2 Flow diagram of statistical analysis in this study

70x61mm (300 x 300 DPI)

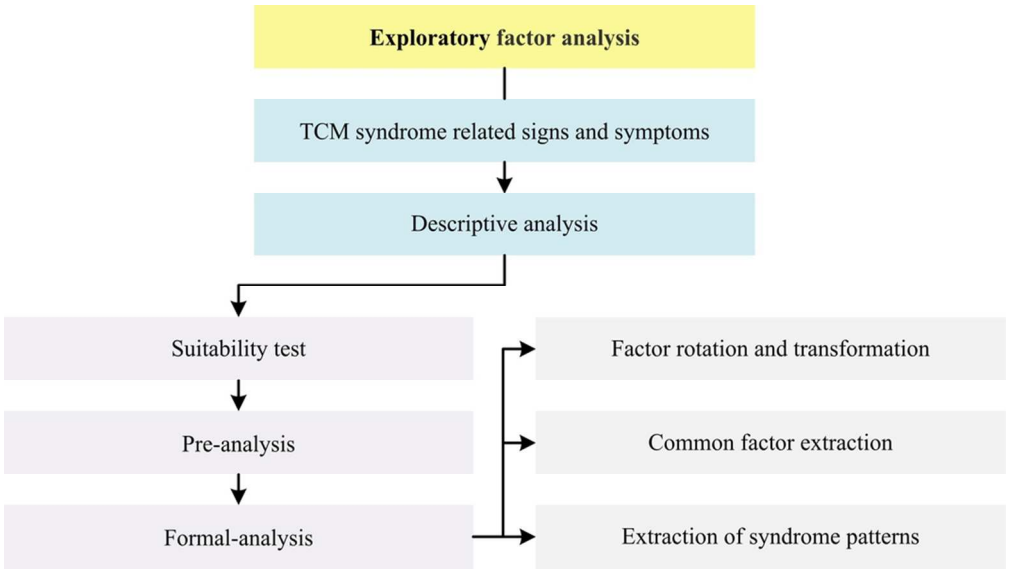


Fig.3 Flow diagram of exploratory factor analysis in this study

89x50mm (300 x 300 DPI)

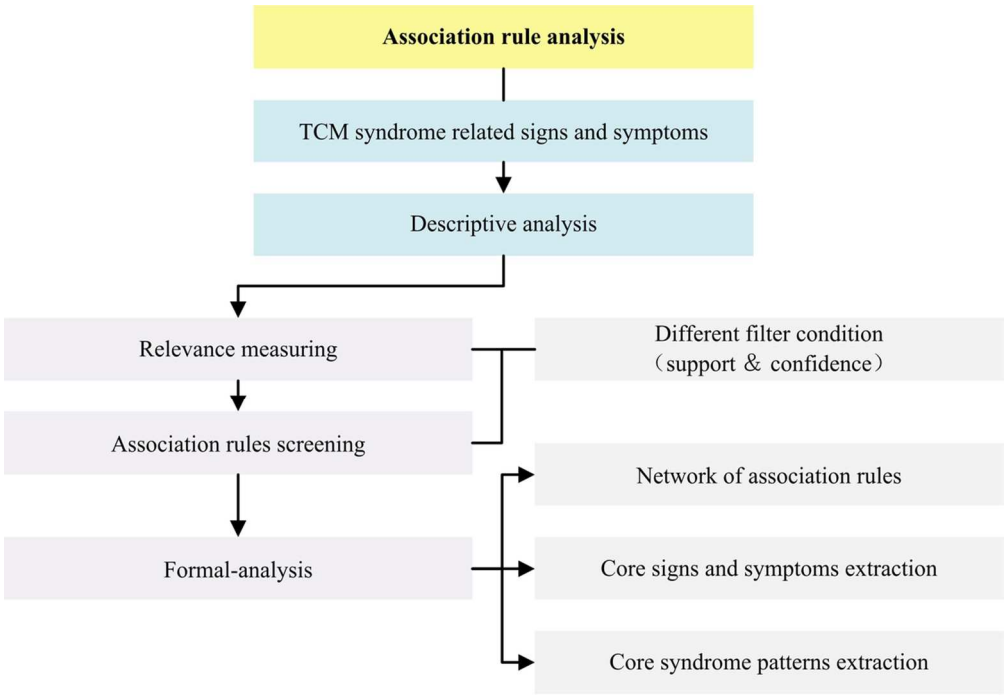


Fig.4 Flow diagram of association rule analysis in this study

110x75mm (300 x 300 DPI)

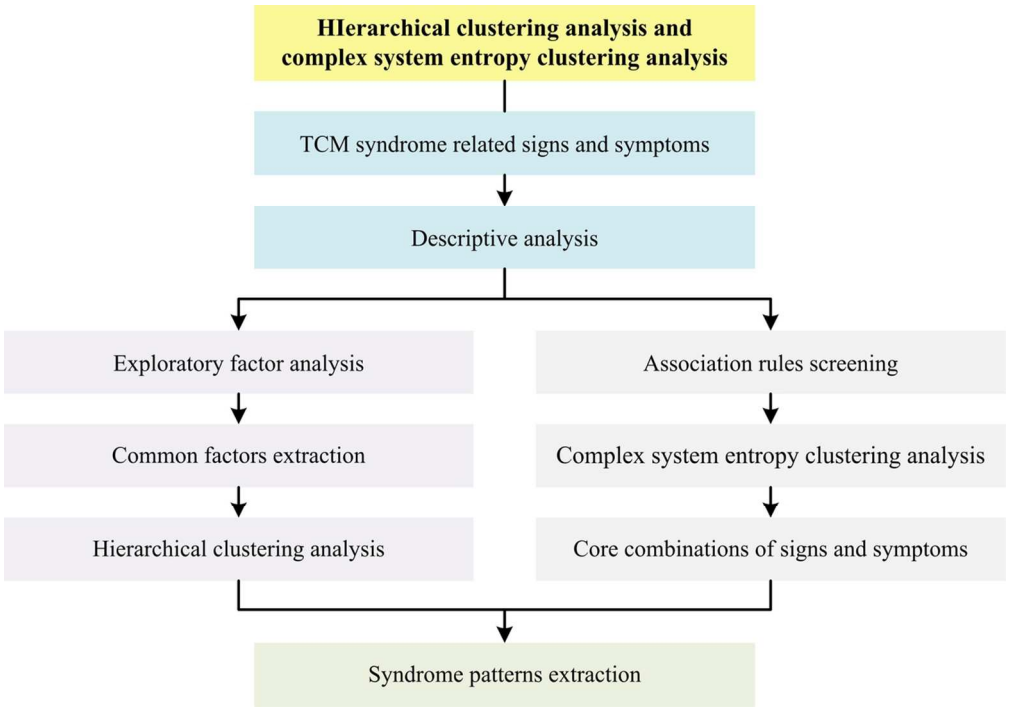


Fig.5 Flow diagram of hierarchical clustering analysis and complex system entropy clustering analysis in this study

112x78mm (300 x 300 DPI)

STROBE 2007 (v4) checklist of items to be included in reports of observational studies in epidemiology*

Checklist for cohort, case-control, and cross-sectional studies (combined)

Section/Topic	Item #	Recommendation	Reported on page #
Title and abstract	1	(*a*) Indicate the study's design with a commonly used term in the title or the abstract	1
		(*b*) Provide in the abstract an informative and balanced summary of what was done and what was found	1
Introduction			
Background/rationale	2	Explain the scientific background and rationale for the investigation being reported	3
Objectives	3	State specific objectives, including any pre-specified hypotheses	3,4
Methods			
Study design	4	Present key elements of study design early in the paper	4
Setting	5	Describe the setting, locations, and relevant dates, including periods of recruitment, exposure, follow-up, and data collection	4,5
Participants	6	(*a*) *Cohort study*—Give the eligibility criteria, and the sources and methods of selection of participants. Describe methods of follow-up *Case-control study*—Give the eligibility criteria, and the sources and methods of case ascertainment and control selection. Give the rationale for the choice of cases and controls *Cross-sectional study*—Give the eligibility criteria, and the sources and methods of selection of participants	5
		(*b*) *Cohort study*—For matched studies, give matching criteria and number of exposed and unexposed *Case-control study*—For matched studies, give matching criteria and the number of controls per case	Not applicable
Variables	7	Clearly define all outcomes, exposures, predictors, potential confounders, and effect modifiers. Give diagnostic criteria, if applicable	5,6
Data sources/ measurement	8*	For each variable of interest, give sources of data and details of methods of assessment (measurement). Describe comparability of assessment methods if there is more than one group	5,6
Bias	9	Describe any efforts to address potential sources of bias	4,5
Study size	10	Explain how the study size was arrived at	4
Quantitative variables	11	Explain how quantitative variables were handled in the analyses. If applicable, describe which groupings were chosen and why	7
Statistical methods	12	(*a*) Describe all statistical methods, including those used to control for confounding	7
		(*b*) Describe any methods used to examine subgroups and interactions	7
		(*c*) Explain how missing data were addressed	7
		(*d*) *Cohort study*—If applicable, explain how loss to follow-up was addressed *Case-control study*—If applicable, explain how matching of cases and controls was addressed	Not applicable

		Cross-sectional study—If applicable, describe analytical methods taking account of sampling strategy	
		(*e*) Describe any sensitivity analyses	Not applicable
Results			
Participants	13*	(a) Report numbers of individuals at each stage of study—eg numbers potentially eligible, examined for eligibility, confirmed eligible, included in the study, completing follow-up, and analysed	Not applicable (protocol)
		(b) Give reasons for non-participation at each stage	Not applicable (protocol)
		(c) Consider use of a flow diagram	4,6
Descriptive data	14*	(a) Give characteristics of study participants (eg demographic, clinical, social) and information on exposures and potential confounders	Not applicable (protocol)
		(b) Indicate number of participants with missing data for each variable of interest	Not applicable (protocol)
		(c) *Cohort study*—Summarise follow-up time (eg, average and total amount)	Not applicable (protocol)
Outcome data	15*	*Cohort study*—Report numbers of outcome events or summary measures over time	Not applicable (protocol)
		Case-control study—Report numbers in each exposure category, or summary measures of exposure	Not applicable (protocol)
		Cross-sectional study—Report numbers of outcome events or summary measures	Not applicable (protocol)
Main results	16	(*a*) Give unadjusted estimates and, if applicable, confounder-adjusted estimates and their precision (eg, 95% confidence interval). Make clear which confounders were adjusted for and why they were included	Not applicable (protocol)
		(*b*) Report category boundaries when continuous variables were categorized	Not applicable (protocol)
		(*c*) If relevant, consider translating estimates of relative risk into absolute risk for a meaningful time period	Not applicable (protocol)
Other analyses	17	Report other analyses done—eg analyses of subgroups and interactions, and sensitivity analyses	Not applicable (protocol)
Discussion			
Key results	18	Summarise key results with reference to study objectives	Not applicable (protocol)

Limitations	19	Discuss limitations of the study, taking into account sources of potential bias or imprecision. Discuss both direction and magnitude of any potential bias	10
Interpretation	20	Give a cautious overall interpretation of results considering objectives, limitations, multiplicity of analyses, results from similar studies, and other relevant evidence	9,10
Generalisability	21	Discuss the generalisability (external validity) of the study results	10
Other information			
Funding	22	Give the source of funding and the role of the funders for the present study and, if applicable, for the original study on which the present article is based	10

*Give information separately for cases and controls in case-control studies and, if applicable, for exposed and unexposed groups in cohort and cross-sectional studies.
Note: An Explanation and Elaboration article discusses each checklist item and gives methodological background and published examples of transparent reporting. The STROBE checklist is best used in conjunction with this article (freely available on the Web sites of PLoS Medicine at http://www.plosmedicine.org/, Annals of Internal Medicine at http://www.annals.org/, and Epidemiology at http://www.epidem.com/). Information on the STROBE Initiative is available at www.strobe-statement.org.

BMJ Open

Validating traditional Chinese syndrome features in varied stages of chronic gastritis malignant transformation: study protocol for a cross-sectional study

Journal:	*BMJ Open*
Manuscript ID	bmjopen-2017-020939.R2
Article Type:	Protocol
Date Submitted by the Author:	02-Jul-2018
Complete List of Authors:	Zhang, Yin; Bregham and Women's Hospital and Harvard Medical School, Medicine; Beijing University of Chinese Medicine, School of Chinese Medicine Liu, Yue; Beijing University of Chinese Medicine, School of Chinese Medicine Song, Rui; The Association of Chinese American Physicians; American Chinese Medical Exchange Society Zhang, Li; China-Japan Friendship Hospital, Department of Neurosurgery Su, Zeqi; Beijing University of Chinese Medicine, School of Chinese Medicine Li, Yannan; Affiliated Hospital of Shandong University of Chinese Medicine, Department of Gynecology Chen, Runhua; Dongfang Hospital Affiliated to Beijing University of Chinese Medicine, Department of Gastroenterology Shi, Ning; Beijing Electric Power Hospital, Department of Health Management Zhao, Xia; Beijing University of Chinese Medicine, School of Humanities Du, Shiyu; China-Japan Friendship Hospital, Department of Gastroenterology Ding, Xia; Beijing University of Chinese Medicine, School of Chinese Medicine
Primary Subject Heading:	Complementary medicine
Secondary Subject Heading:	Gastroenterology and hepatology
Keywords:	Chronic gastritis, Malignant transformation, Traditional Chinese medicine, Syndrome, Cross-sectional study

Validating traditional Chinese syndrome features in varied stages of chronic gastritis malignant transformation: study protocol for a cross-sectional study

Yin Zhang,[1,2,3,a] Yue Liu,[1,a] Rui Song,[4,5,a] Li Zhang,[6] Zeqi Su,[1] Yannan Li,[7] Runhua Chen,[8,1] Ning Shi,[9,1] Xia Zhao,[10] Shiyu Du,[11] and Xia Ding [1,*]

[1]School of Chinese Medicine, Beijing University of Chinese Medicine, Chaoyang District, Beijing 100029, China

[2]Channing Division of Network Medicine, Department of Medicine, Brigham and Women's Hospital, Harvard Medical School, Boston, MA 02115, USA

[3]Institute of Basic Research in Clinical Medicine, China Academy of Chinese Medical Sciences, Dongcheng District, Beijing 100700, China

[4]The Association of Chinese American Physicians, Flushing, NY 11354, USA

[5]American Chinese Medical Exchange Society, Burlington, MA 01803, USA

[6]Department of Neurosurgery, China-Japan Friendship Hospital, Chaoyang District, Beijing 100029, China

[7]Department of Gynecology, Affiliated Hospital of Shandong University of Chinese Medicine, Lixia District, Jinan 250011, China

[8]Department of Gastroenterology, Dongfang Hospital Affiliated to Beijing University of Chinese Medicine, Beijing 100078, China

[9]Department of Health Management, Beijing Electric Power Hospital, Fengtai District, Beijing 100073, China

[10]School of Humanities, Beijing University of Chinese Medicine, Chaoyang District, Beijing 100029, China

[11]Department of Gastroenterology, China-Japan Friendship Hospital, Chaoyang District, Beijing 100029, China

[a] These authors contributed equally to this work.

[*] Correspondence to: Dr. Xia Ding, E-mail: dingx@bucm.edu.cn

Word count: 2900 words (excluding title page, abstract, references, figures and tables)

Abstract

Introduction

The transition from chronic non-atrophic gastritis (CNAG) to chronic atrophic gastritis (CAG) and gastric carcinoma (GC) is regarded as a representative disease model of gastric mucosa malignant transformation led by uncontrolled inflammation. Traditional Chinese medicine (TCM) syndrome-targeted therapies have been applied in treating chronic gastritis (CG) malignant transformation in China with satisfying efficacy. This study aims to validate TCM syndrome features in each stage of CG malignant transformation. The findings may shed light on the TCM hypothesis of CG malignant transformation, and thus optimize syndrome-targeted treatment strategies of CNAG, CAG, and GC respectively.

Methods and analysis

The present study is a cross-sectional study conducted in China. 2000 eligible patients, including 500 CNAG cases, 1000 CAG cases, and 500 GC cases, will be recruited from 4 TCM hospitals. Primary outcome measures include the prevalence of TCM syndrome patterns in varied stages of CG malignant transformation. Secondary outcome measures include prevalence and severity of all the presenting signs and symptoms collected by using TCM four diagnostic methods. Descriptive analysis, comparative analysis, and correlation analysis of all the measurement data will be performed by biostatisticians. Unsupervised data mining analyses, including exploratory factor analysis, association rule analysis, hierarchical clustering analysis, and complex system entropy clustering analysis, etc., will also be performed by data scientists respectively for in-depth analyses of TCM syndrome-related indicators.

Ethics and dissemination

The protocol has been approved by Ethical Review Board of Donzhimen Hospital Affiliated to Beijing University of Chinese Medicine (No. ECPJ-BDY-2014-02). All the study outcomes will be disseminated through national conference reports and in the meantime published on peer-reviewed journals.

Study registration

This study has been registered on ClinicalTrials.gov (No. NCT03314038).

Keywords:

Chronic gastritis; Malignant transformation; Traditional Chinese medicine; Syndrome; Cross-sectional study

Strengths and limitations of this study

►To the authors' knowledge, this is among the first reported large population based studies on validating TCM syndrome features in varied stages of CG malignant transformation.

►A large sample size of 2000 participants can be achieved. Full-scale TCM and modern medicine indicators can be accurately collected.

►Both statistical analysis and unsupervised data mining strategies will be applied by research personnel parallelly to reinforce complementary strengths.

►This is not a prospective study, and thus longitudinal data couldn't be collected by research personnel to speculate causal relationship.

►Selection bias and information bias may exist.

Introduction

Chronic gastritis (CG), persistent inflammation of the gastric mucosa with no specific clinical manifestations, has been regarded as the most common gastrointestinal disease with an incidence of 70% among the adult population in China [1]. Chronicity and recurrence of chronic non-atrophic gastritis (CNAG) may lead to chronic atrophic gastritis (CAG), which is characterized by gastric mucosal atrophy and usually accompanied by metaplasia or dysplasia [2]. As the foremost step of CG malignant transformation, CAG is acknowledged as premalignant lesion of gastric cancer (GC). The sequential stages of inflammation, metaplasia, dysplasia and carcinoma have demonstrated a well-established tumorigenesis model of intestinal type of GC leaded by uncontrolled inflammation [2-4]. Statistics have shown that there are 951,600 annual new cases of GC worldwide [5]. China has been reported as one of the countries with largest population of both CG and GC sufferers [1]. Considering the severe burden of above-mentioned diseases, active intervention on CG malignant transformation process is highly needed [6-8].

Traditional Chinese medicine (TCM), as one of the most important and time-honored alternative medicine approaches worldwide, has developed unique theories of etiology and diagnosis since ancient time. According to TCM theory, syndrome is defined as a categorised pattern of symptoms and signs in a patient at a specific stage during the course of specific diseases, and is considered the most important unit for evaluating pathogenesis. TCM practitioners have been applying the central principles of syndrome pattern differentiation in clinical practice to identify the physical condition of patients and the pathologensis of diseases with a history of more than 2500 years. Nowadays, TCM physicians keep on using the prestigious and traditional "four diagnostic methods" of looking, listening/smelling, asking and palpating to collect the information of symptoms and signs on patients comprehensively. Full-scale assessment of aforementioned clinical information can then be conducted by physicians based on TCM theory and clinical experience to reveal the etiology (syndrome) of diseases, and thus guiding individualized herbal prescriptions. Though huge leaps have been seen in past decades in elaborating the pathogenesis of CAG, modern medicines remain unsatisfied in treating CNAG and CAG to some extent [9]. Owing to the heavy burden of recurrence, chronicity, and malignant transformation of CG, a considerable proportion of sufferers in China have focused on TCM therapies.

Trials have indicated the benefits of TCM syndrome-targeted interventions on CNAG and CAG [10-14]. The Society of Gastroenterology, China Association of Chinese Medicine has propagated the latest "*TCM Consensus on Chronic Superficial Gastritis Diagnosis and Treatment*" and "*TCM Consensus on Chronic Atrophic Gastritis Diagnosis and Treatment*" respectively in 2009 [15, 16]. The Society of Gastroenterology, China Society of Integrated Traditional Chinese and Western Medicine also issued "*Integrative Medicine Consensus on Chronic Gastritis Diagnosis and Treatment in China*" in 2012 [17]. The aforementioned consensuses emphasized the importance of applying TCM syndrome-targeted therapies in controlling gastric mucosal inflammation and atrophy, and inhibiting metaplasia and dysplasia in accordance with China's local conditions. "*Consensus on chronic gastritis in China*" promulgated by the Society of Gastroenterology, Chinese Medical Association and "*Diagnosis and Treatment Guideline for Gastric Carcinoma*" issued by the National Health and Family Planning Commission of the People's Republic of China also recommended the application of TCM syndrome differentiation based therapies in addition to conventional treatment respectively [1, 18].

Both consensus and literature reviews have also indicated the potential links between syndrome patterns and CG malignant transformation [15-17, 19-21]. We believe that the diversity and dynamics of

TCM syndromes may play a role all the way through this course. However, no large population based study has been reported to support in-depth analysis of prevalence and severity of varied TCM syndromes in different stages of CG malignant transformation. Taking note of the limitations of previous studies, we highlight validating detailed characteristics of TCM syndromes in CNAG, CAG, and GC population respectively in the present study. The new findings may shed light on the TCM pathogenesis theory of CG malignant transformation, and thus optimize syndrome-targeted treatment strategies by TCM herbal prescriptions in distinct stages of this process.

Study aims

The primary goal for the present study is to identify TCM syndrome features in different stages of CG malignant transformation respectively, including CNAG, CAG, and GC. In addition, this study can also be regarded as a pilot study speculating potential dynamic features of TCM syndrome in the whole course of CG. These features may provide important clues for the design of future observational long-term follow-up studies aiming to further explore novel risk prediction and syndrome evolution models combining both TCM and modern medicine indicators for CG or CAG malignant transformation.

Methods and analysis

Design

This is a cross-sectional study undertaken from 2014 to 2018, comprising 2000 eligible participants in 4 hospitals in China. The general flow diagram of this study is outlined in Fig. 1.

Fig.1 General flow diagram of this study

Recruitment

This study is led by Beijing University of Chinese Medicine (BUCM). The 4 medical centres participating in the present study include Dongzhimen Hospital Affiliated to BUCM, Dongfang Hospital Affiliated to BUCM, the 3rd Affiliated Hospital of BUCM, and Wangjing Hospital Affiliated to China Academy of Chinese Medical Sciences (CACMS). To make a definitive diagnosis for each participant, an upper gastrointestinal endoscopy based screening process is conducted by study personnel during recruitment. All the eligible patients presenting to the the aforementioned hospitals are then identified and recruited continuously until the required sample size for each category (CNAG, CAG, and GC) is achieved respectively. This study has opened to recruitment in 2014. Recruitment is ongoing and is expected to conclude by the end of 2018. The sample size required is calculated using the following formula: $N=(Z^2 p(1-p))/d^2$. N is sample size; P is expected proportion of certain syndrome, if 50%, P=0.5; d is precision, if 5%, d=0.05; Z is the level of confidence of 95%, conventionally, Z value is 1.96 for 95% CI. According to this formaula, the sample size would be 385 cases for each category (CNAG, CAG, and GC). However, taking into full consideration of items of case report forms (CRF), experts' opinions, study expenses, and operability, the anticipated sample size is finally decided to be 2000 in total, including 500 CNAG cases, 1000 CAG cases, and 500 GC cases.

Diagnostic criteria

Diagnostic criteria of CNAG, CAG, GMA, IM, and GED in the present study refer to "Consensus on chronic gastritis in China" promulgated by the Society of Gastroenterology, Chinese Medical Association

[1]. Diagnostic criteria of GC refers to "*Diagnosis and Treatment Guideline for Gastric Carcinoma (trail version)*" issued by the National Health and Family Planning Commission of the People's Republic of China [18]. Diagnostic criteria of Helicobacter pylori (Hp) infection refers to "*Chinese Consensus on Helicobacter pylori Infection Treatment*" issued by the Chinese Society of Gastroenterology, Chinese Medical Association [22]. Criteria for TCM syndrome pattern differentiation diagnosis in this study refers to "*TCM Consensus on Chronic Superficial Gastritis Diagnosis and Treatment*" [15] and "*TCM Consensus on Chronic Atrophic Gastritis Diagnosis and Treatment*" [16] issued by the Society of Gastroenterology, China Association of Chinese Medicine.

Inclusion criteria

The inclusion criteria are as follows: (1) Meeting diagnostic criteria of CNAG, CAG, or GC; (2) Willing to cooperate with investigators for data and tissue sample collection; (3) Willing to sign informed consent.

Exclusion criteria

The exclusion criteria are as follows: (1) History of previous stomach surgeries; (2) Unable to participate in data and sample collection for any reason.

Data and samples collection

Strict training in standard operating procedure (SOP) of clinical data and samples collection for research personnel is conducted by principal investigator before study initiation. Prior to recruitment, participants are required to sign informed consents. In this process, clinial investigators take the responsibility for explaining the objectives, general procedures, data collection methods, risks and benefits, authorization, data privacy policy, and other necessary details to every patient. At the time of the data collection the investigators are blinded for the distinct groups that patients may be included. Research personnel will then assess all the indicators and tissue samples needed at once during recruitment. Each participant is interviewed by at least two physicians to ensure quality control. A list of degree-definition of all the TCM indicators is attached to each CRF for investigators and participants to refer to.

Outcome measures and overview of CRFs

Primary outcome measures of this study include the prevalence of TCM syndrome patterns in varied stages of CG malignant transformation (CNAG, CAG, and GC). Secondary outcome measures include prevalence and severity of all the presenting signs and symptoms collected by using TCM four diagnostic methods (looking, listening/smelling, asking, and palpating). The aforementioned clinical features are identified in detail base on TCM syndrome pattern differentiation scales.

Accurate modern medicine diagnostic information in this study is achieved based on latest gastroscope reports, and pathological reports. Hp infects more than half of the populations worldwide and causes long-term progressive damage to the gastric mucosa in all infected individuals, is acknowledged as one of the most important risk factors for CG and GC [23-26]. Stage-specific dietary intake, ingestion of ascorbic acid and nitrate, and other factors have also been proved to associate with the multistep and multifactorial process of gastric carcinogenesis [1-4, 27]. Considering the importance of above-mentioned indicators, full-scale information regarding to Hp test reports, dietary intake, lifestyle, medication, etc. are collected in addition to TCM indicators. Other necessary information collected in our study covers demographic

information, disease history, complications, and laboratory indices. In addition, tissue samples collection is also included. Subgroup analysis can thus be carried out in a number of specific populations (e.g. CAG with metaplasia and/or dysplasia, CAG with/without Hp infection, etc.).

According to study design and CRFs, detailed content of data and tissue samples collection are illustrated in Table. 1.

Table.1 Content of data and tissue samples collection

Content	Details
Demographic Information	Name, date of birth, gender, age, nationality, height, weight, level of education, etc.
Diagnostic Information	Latest gastroscope reports, Hp test reports, and pathological reports
TCM Syndrome Related Information	Features of presenting signs and symptoms collected by using TCM four diagnostic methods of looking, listening/smelling, asking, and palpating
	TCM syndrome pattern differentiation scale scores
	Results of TCM syndrome identification
Disease History	Past history and family history of CG, GC, and Hp infection, etc.
Complications	Upper gastrointestinal hemorrhage, upper gastrointestinal ulcer, pyloric obstruction, anemia, gastric carcinoma metastasis, etc.
Dietary Intake and Lifestyle	Dietary structure, working status, psychological status, history of smoking and alcohol intake, etc.
Medication	Current drug use on CG and GC treatment (current use of proton pump inhibitors, etc. at the time of upper endoscopy), and nonsteroidal anti-inflammatory drugs (NSAIDs)
Laboratory Indices	Routine blood test, routine stool test, tumor markers, etc.
Tissue Samples	Tissue samples of diseased gastric mucosa

Patient and public involvement

Before the formal recruitment started, a pre-test of CRF was conducted by research personnel in patients at the Dongzhimen Hospital Affiliated to BUCM. In this process, principal investigator and research personnel collated the feedback from the patients and using it to improve the final design of the CRF. Study outcomes will be disseminated on conference reports and academic publications. Main findings will also be disseminated to study participants via email and the BUCM website.

Data management

SOP of data collection, entry, editing, locking and retrieval is set up by data management centre of this study in Beijing University of Chinese Medicine. Confidentiality, authenticity, and integrity of all the clinical information gathered in the present study is highlighted and maintained at all levels of data management. CRFs are organized by unique identification numbers and thus no patients' personal

identifiers will be available in any records. Before data-entry, double-checking of range and logic of every variable value is performed by primary researchers. Data is then double-entered into a dedicated database by data administrators to confirm there is no input error.

Data analysis and outcome reporting

In order to achieve the primary goal of validating TCM syndrome features in varied stages of CG malignant transformation, varied data analysis strategies on multidimensional indicators will be performed parallelly from both biostatistics and data mining perspectives.

Descriptive statistics will be carried out by biostatisticians for all the measurement data, including demographic and clinical characteristics in the whole study population and each stratum. Comparative analysis of qualitative TCM and modern medicine indicators in different groups will be performed using chi-squared tests. Comparative analysis of quantitative scores of TCM syndrome pattern differentiation scales will be conducted using rank-sum test. A probability of $P < 0.05$ will be considered statistically significant for these analyses. Correlation analysis of pathological grades and stages of CG malignant transformation and TCM indicators will be performed using logistic regression model. The significance level for introducing and removing variables will be 0.05 and 0.10.

In addition, unsupervised data mining analyses, including exploratory factor analysis, association rule analysis, hierarchical clustering analysis, and complex system entropy clustering analysis, etc., will be performed by data scientists respectively for indepth exploring of TCM syndrome distribution and evolution features in the whole course of CG malignant transformation.

Statistical analyses in this study will be carried out using SAS software *(version 9.3, SAS Institute Inc., Cary, NC, U.S.A)* , and R software *(version 3.3.1, The R Project for Statistical Computing., Auckland, N.Z.)*. Data mining analyses in this study will be processed by SPSS Clementine software *(version 12.0, SPSS Inc., Chicago, IL, U.S.A)*, and Matlab Software *(version 2016a, Mathworks Inc., Natick, MA, U.S.A)*. Plotting in this study will be performed using Office Excel software *(version 2007, Microsoft Inc., Redmond, WA, U.S.A)* and Office Visio software *(version 2007, Microsoft Inc., Redmond, WA, U.S.A)*.

General flow diagram of biostatistical ananlysis process is outlined in Fig. 2. Detailed flow diagrams of aforementioned exploratory data mining strategies to be conducted in this study are illustrated respectively in Fig. 3-5.

Fig.2 Flow diagram of statistical analysis strategy in this study

Fig.3 Flow diagram of exploratory factor analysis in this study

Fig.4 Flow diagram of association rule analysis in this study

Fig.5 Flow diagram of hierarchical and complex system entropy clustering analysis in this study

Study registration

This study has been registered on ClinicalTrials.gov (No. NCT03314038).

Ethics and Dissemination

The protocol is designed in accordance with the Biological Ethics Review Method Involving Humans by the Ministry of Health of the People's Republic of China, and the Declaration of Helsinki. Ethical Review Board of Donzhimen Hospital Affiliated to Beijing University of Chinese Medicine has approved the protocol (No. ECPJ-BDY-2014-02). All the information will be used for only academic aims.

All the study outcomes will be disseminated on national conference reports, peer-reviewed TCM and modern medicine journals, or other academic publications. Findings from this study will not only provide much-needed understandings in this topic for both TCM clinical practitioners and researchers, but also greatly contribute to optimization of further prospective study design.

Strengths and limitations of this study

Strengths of our study should be addressed. Firstly, the current large cross-sectional study is the first research conducted to reveal TCM syndrome features in varied stages of CG malignant transformation. Although the current study design may limit generalisability of the findings, considering the expenses of prospective study and the relatively low yearly malignant transformation rate of CG individuals, it is important for us to carry out this pilot study speculating potential TCM syndrome dynamic features from CG to GC. The study outcomes will provide targeted clues for the future design of long-term follow-up studies aiming to further explore novel risk prediction and TCM syndrome evolution models for CNAG or CAG malignant transformation combining both TCM and modern medicine indicators. Secondly, diagnosis of CNAG, CAG, GC, IM, GED and Hp infection in this study is confirmed based on latest gastroscope reports, Hp test reports, and pathological reports at the time of recruitment. The accurate diagnostic information collected will greatly support conducting detailed subgroup analysis stratified by specific stages and pathological conditions (e.g., CAG with IM and/or GED, CAG with/without Hp infection, etc.). Thirdly, full-scale information collected in this study provides a good basis for in-depth exploration of clinical features. In addition, the parallelly applied biostatistical analysis strategies and unsupervised data mining strategies can reinforce complementary strengths and help us achieve a better understanding of TCM syndrome features in varied stages of these diseases.

Our study also has several limitations. Firstly, this cross-sectional study merely involves data collection at a defined time (recruitment), dynamic characteristics of each invidual in the whole course of disease could thus not be tracked. Further large prospective studies with long follow-up time are highly demanded to expand findings of this study in a broaden range of settings, and thus contribute to building longitudinal data based novel risk assessment models of CG malignant transformation combining both TCM and modern medicine indicators. Secondly, as data are collected merely from TCM hospitals in Beijing, the current participants may not be representative enough to show the clinical features of patients from modern medicine hospitals, or medical centres in other regions of China. Selection bias may thus exist. Thirdly, some of the signs and symptoms related TCM indicators may be complicated enough for patients to fully understand and correctly report without prior TCM knowledge. A list of degree-definition of all the TCM indicators has been attached to each CRF for investigators and participants to refer to in recruitment, and each patient is interviewed by at least two TCM physicians to ensure quality control in this study.

However, information bias may still inevitably exist to some extent, especially due to the complexity of individual self-reported TCM indicators.

Acknowledgments

The authors assume full responsibility for study design, programming, data analyses and interpretation of the outcomes. All authors approved the final version of the manuscript. We are grateful to all the participants and staffs from Dongzhimen Hospital Affiliated to BUCM, Dongfang Hospital Affiliated to BUCM, Wangjing Hospital Affiliated to CACMS, and the 3rd Affiliated Hospital of BUCM for their remarkable contribution to this study.

Authors' contributions

Study concept: Xia Ding

Study and statistical analysis design: Xia Ding, Yin Zhang, Yue Liu, Rui Song, and Zeqi Su

Drafting of the manuscript: Yin Zhang, Yue Liu, and Rui Song

Critical review and revisions of the manuscript: Xia Ding, Li Zhang, Yannan Li, Runhua Chen, Ning Shi, Xia Zhao, and Shiyu Du

Yin Zhang, Yue Liu and Rui Song contributed equally to this work.

Funding statement

This work was supported by the National Natural Science Foundation of China (No. 81630080, 91129714, 81270466, 30500183, 30870990, 81173424, 81373796, 81673793), National Undergraduates Innovating Experimentation Project of the Ministry of Education of China (No. 081002609), Research Foundation of the Doctoral Program of Higher Education of China (No. 20120013110014), and Science Research Foundation of BUCM (No. 2014-JYBZZ-XS-134).

Competing interest statement

The authors report no competing interests.

Abbreviations

CG: Chronic gastritis; CNAG: Chronic non-atrophic gastritis; CAG: Chronic atrophic gastritis; GC: Gastric carcinoma; TCM: Traditional Chinese medicine.

References

1. Fang YJ, Liu WZ, Li ZS, *et al*. Consensus on chronic gastritis in China (version 2012). *Chin J Gastroenterol* 2013;18:24-36. [Chinese]
2. Dinis-Ribeiro M, Areia M, De Vries AC, *et al*. Management of precancerous conditions and lesions in the stomach (MAPS): guideline from the European Society of Gastrointestinal Endoscopy (ESGE), European Helicobacter Study Group (EHSG), European Society of Pathology (ESP), and the Sociedade Portuguesa de Endoscopia Digestiva (SPED). *Endoscopy* 2012;44:74-94.
3. Correa P. A human model of gastric carcinogenesis. *Cancer Res* 1988;48:3554–60,.
4. Correa P. Human gastric carcinogenesis: a multistep and multifactorial process - First American Cancer Society Award Lecture on Cancer Epidemiology and Prevention. *Cancer Res* 1992;52:6735-40.

5. Torre LA, Bray F, Siegel RL, *et al*. Global cancer statistics, 2012. *CA-Cancer J Clin* 2012;65:87-108.
6. Park YH, Kim N. Review of atrophic gastritis and intestinal metaplasia as a premalignant lesion of gastric cancer. *J Cancer Prev* 2015;20:25-40.
7. Correa P, Piazuelo MB. The gastric precancerous cascade. *J Digest Dis* 2012;13:2-9.
8. De Vries AC, Van Grieken NC, Looman CW, *et al*. Gastric cancer risk in patients with premalignant gastric lesions: a Nationwide Cohort Study in the Netherlands. *Gastroenterology* 2008;134:945-52,.
9. Den Hollander WJ, Kuipers EJ. Current pharmacotherapy options for gastritis. *Expert Opin Pharmaco* 2012;13:2625-36.
10. Dai YK, Zhang YZ, Li DY, *et al*. The efficacy of Jianpi Yiqi therapy for chronic atrophic gastritis: a systematic review and meta-analysis. *PLoS One* 2017;12:e0181906.
11. Cheng RD, Zhao B, Ma SQ, *et al*. Therapeutic effect of activating blood therapy in treating chronic atrophic gastritis: a meta-analysis. *Chin J Tradit Chin Med Pharm* 2017;32:2050-4. [Chinese]
12. Guo ZL, Su ZN, Wang ZF, *et al*. Therapeutic effect of modified Sijunzi Tang in treating chronic atrophic gastritis: a meta-analysis," *Chin J Exp Tradit Med* 2015;21:204-8. [Chinese]
13. Yang JX, An J, Peng JS, *et al*. Therapeutic effect of Banxia Xiexin decoction in treating chronic atrophic gastritis: a meta-analysis. *J Beijing Univ Chin Med* 2015;38:46-52. [Chinese]
14. Liao WL, Chen GZ, Hu H, *et al*. Therapeutic effect of Weifuchun in treating chronic atrophic gastritis: a meta-analysis. *Liaoning J Tradit Chin Med* 2015;42:11-15. [Chinese]
15. Zhang SS, Li QG, Hang HP, *et al*. Traditional Chinese medicine consensus on chronic superficial gastritis diagnosis and treatment (version 2009). *Chin J Integr Gastroenterol* 2010; 18:207-9. [Chinese]
16. Zhang SS, Li QG, Tang XD, *et al*. Traditional Chinese medicine consensus on chronic atrophic gastritis diagnosis and treatment (version 2009). *Chin J Integr Gastroenterol* 2010;18:345-9. [Chinese]
17. Zhang WD, Li JX, Chen ZS, *et al*. Integrative medicine consensus on chronic gastritis diagnosis and treatment in China (version 2012). *Chin J Integr Tradit West Med* 2012;32:738-43. [Chinese]
18. National Health and Family Planning Commission of the People's Republic of China. "Diagnosis and treatment guideline for gastric carcinoma (trial version). Available from URL: http://www.moh.gov.cn/ewebeditor/uploadfile/2013/05/20130515115755662.pdf
19. Su ZQ, Li PC, Guo Q, *et al*. Syndrome pattern transformation in chronic gastritis. *J Beijing Univ Tradit Chin Med* 2015;38:762-71. [Chinese]
20. Gan HJ, Liang LL, Han ML, *et al*. Relationship between syndrome distribution and chronic atrophic gastritis. *Chin J Tradit Chin Med Pharm* 2015;30:2456-9. [Chinese]
21. Li PC, Ding X, Guo Q, *et al*. Literature review of Chinese medicine syndromes and syndrome element distribution features of chronic gastritis. *Liaoning J Tradit Chin Med* 2014;41:2532-34. [Chinese]
22. Liu WJ, Xie Y, Cheng H, *et al*. Chinese consensus on Helicobacter pylori infection treatment (version 2012). *Chin J Gastroenterol* 2014;17:3-21. [Chinese]
23. Chen Q, Lu H. Kyoto global consensus report on Helicobacter pylori gastritis and its impact on Chinese clinical practice. *J Digest Dis* 2016;17:353-6.

24. Sugano K, Tack J, Kuipers EJ, *et al*. Kyoto global consensus report on Helicobacter pylori gastritis. *Gut* 2015;64:1353-67.

25. Malfertheiner P, Mégraud F, O'Morain CA, *et al*. Management of Helicobacter pylori infection - the Maastricht IV Florence Consensus Report. *Gut* 2012;61:646–64.

26. World Gastroenterology Organisation. World Gastroenterology Organization Global Guideline: Helicobacter pylori in developing countries. *J Clin Gastroenterol* 2011;45:383-8.

27. Hsiung HY, Fann JC, Yen AM, *et al*. Stage-specific dietary factors associated with the Correa multistep and multifactorial process of human gastric carcinogenesis. *Nutr Cancer* 2016;68:598-610.

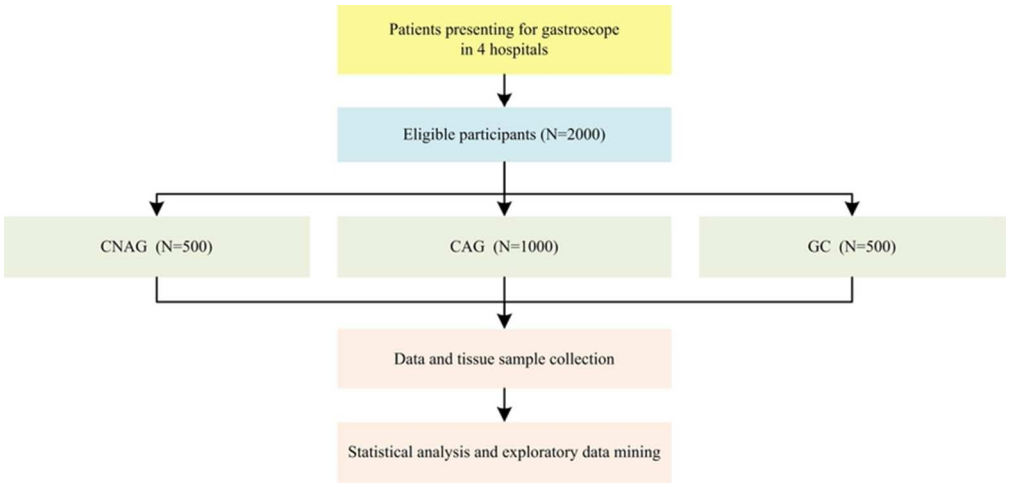


Fig.1 General flow diagram of this study

71x34mm (300 x 300 DPI)

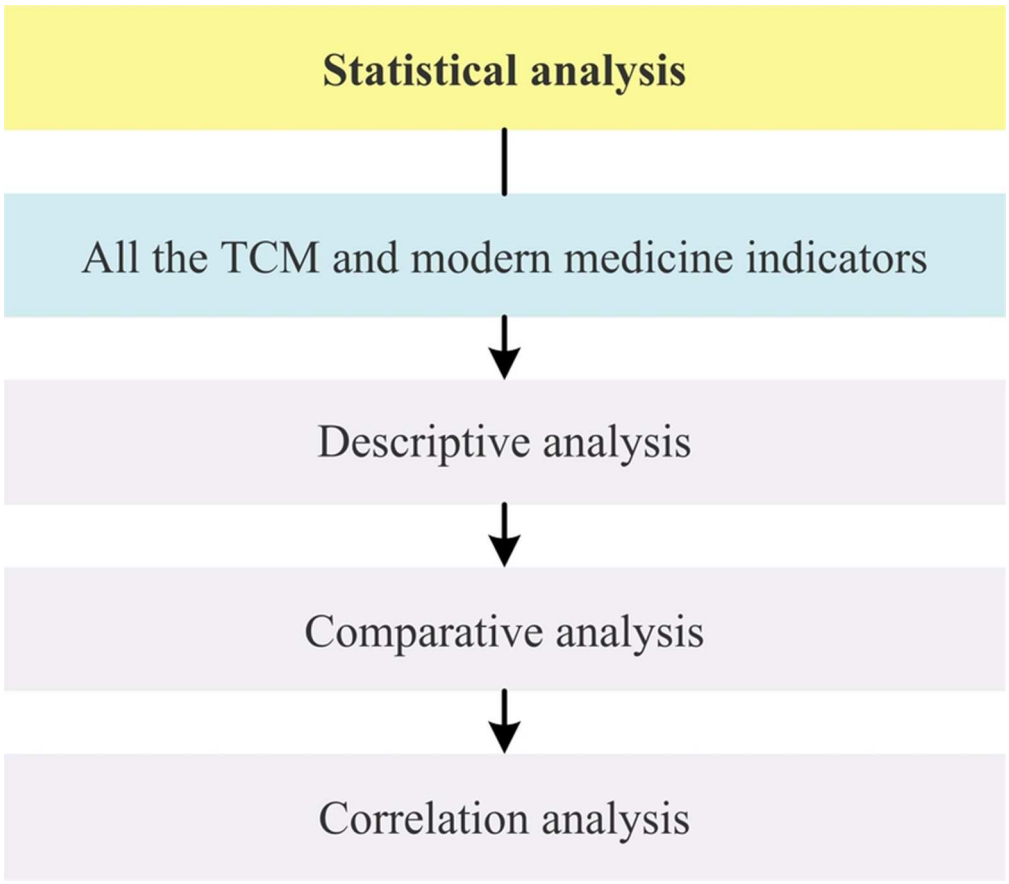


Fig.2 Flow diagram of statistical analysis in this study

70x61mm (300 x 300 DPI)

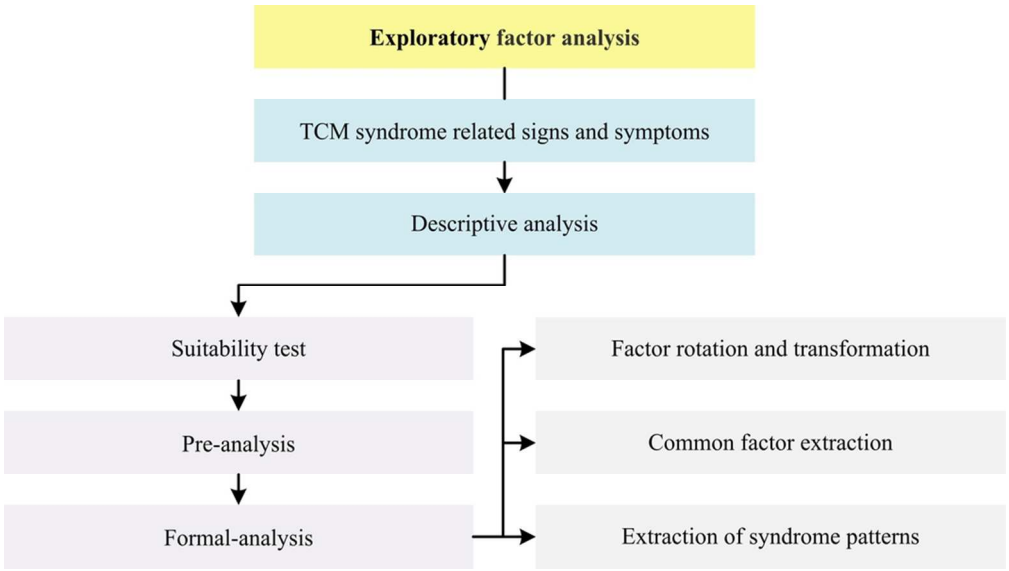


Fig.3 Flow diagram of exploratory factor analysis in this study

89x50mm (300 x 300 DPI)

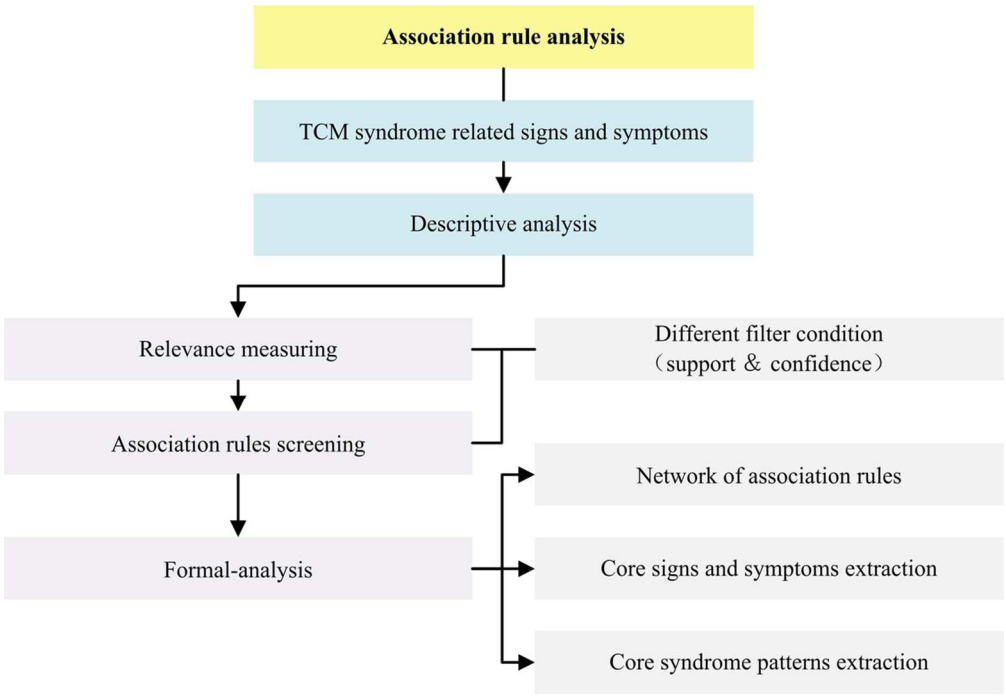


Fig.4 Flow diagram of association rule analysis in this study

110x75mm (300 x 300 DPI)

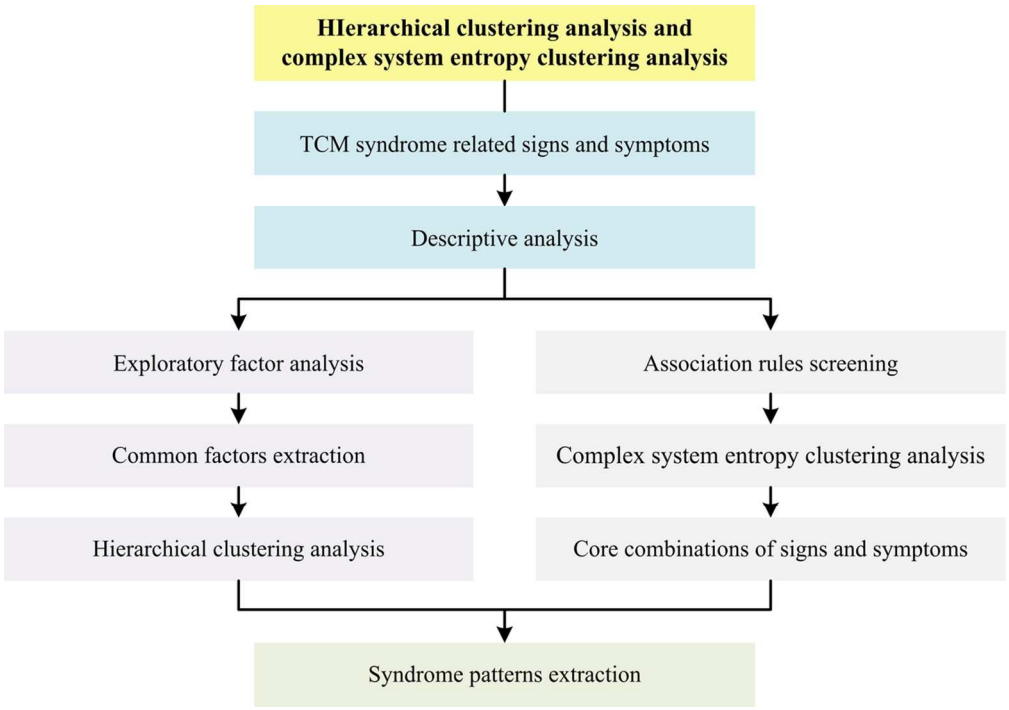


Fig.5 Flow diagram of hierarchical clustering analysis and complex system entropy clustering analysis in this study

112x78mm (300 x 300 DPI)

STROBE 2007 (v4) checklist of items to be included in reports of observational studies in epidemiology*

Checklist for cohort, case-control, and cross-sectional studies (combined)

Section/Topic	Item #	Recommendation	Reported on page #
Title and abstract	1	(*a*) Indicate the study's design with a commonly used term in the title or the abstract	1
		(*b*) Provide in the abstract an informative and balanced summary of what was done and what was found	1
Introduction			
Background/rationale	2	Explain the scientific background and rationale for the investigation being reported	3
Objectives	3	State specific objectives, including any pre-specified hypotheses	3,4
Methods			
Study design	4	Present key elements of study design early in the paper	4
Setting	5	Describe the setting, locations, and relevant dates, including periods of recruitment, exposure, follow-up, and data collection	4,5
Participants	6	(*a*) *Cohort study*—Give the eligibility criteria, and the sources and methods of selection of participants. Describe methods of follow-up *Case-control study*—Give the eligibility criteria, and the sources and methods of case ascertainment and control selection. Give the rationale for the choice of cases and controls *Cross-sectional study*—Give the eligibility criteria, and the sources and methods of selection of participants	5
		(*b*) *Cohort study*—For matched studies, give matching criteria and number of exposed and unexposed *Case-control study*—For matched studies, give matching criteria and the number of controls per case	Not applicable
Variables	7	Clearly define all outcomes, exposures, predictors, potential confounders, and effect modifiers. Give diagnostic criteria, if applicable	5,6
Data sources/ measurement	8*	For each variable of interest, give sources of data and details of methods of assessment (measurement). Describe comparability of assessment methods if there is more than one group	5,6
Bias	9	Describe any efforts to address potential sources of bias	4,5
Study size	10	Explain how the study size was arrived at	4
Quantitative variables	11	Explain how quantitative variables were handled in the analyses. If applicable, describe which groupings were chosen and why	7
Statistical methods	12	(*a*) Describe all statistical methods, including those used to control for confounding	7
		(*b*) Describe any methods used to examine subgroups and interactions	7
		(*c*) Explain how missing data were addressed	7
		(*d*) *Cohort study*—If applicable, explain how loss to follow-up was addressed *Case-control study*—If applicable, explain how matching of cases and controls was addressed	Not applicable

		Cross-sectional study—If applicable, describe analytical methods taking account of sampling strategy	
		(*e*) Describe any sensitivity analyses	Not applicable
Results			
Participants	13*	(a) Report numbers of individuals at each stage of study—eg numbers potentially eligible, examined for eligibility, confirmed eligible, included in the study, completing follow-up, and analysed	Not applicable (protocol)
		(b) Give reasons for non-participation at each stage	Not applicable (protocol)
		(c) Consider use of a flow diagram	4,6
Descriptive data	14*	(a) Give characteristics of study participants (eg demographic, clinical, social) and information on exposures and potential confounders	Not applicable (protocol)
		(b) Indicate number of participants with missing data for each variable of interest	Not applicable (protocol)
		(c) *Cohort study*—Summarise follow-up time (eg, average and total amount)	Not applicable (protocol)
Outcome data	15*	*Cohort study*—Report numbers of outcome events or summary measures over time	Not applicable (protocol)
		Case-control study—Report numbers in each exposure category, or summary measures of exposure	Not applicable (protocol)
		Cross-sectional study—Report numbers of outcome events or summary measures	Not applicable (protocol)
Main results	16	(*a*) Give unadjusted estimates and, if applicable, confounder-adjusted estimates and their precision (eg, 95% confidence interval). Make clear which confounders were adjusted for and why they were included	Not applicable (protocol)
		(*b*) Report category boundaries when continuous variables were categorized	Not applicable (protocol)
		(*c*) If relevant, consider translating estimates of relative risk into absolute risk for a meaningful time period	Not applicable (protocol)
Other analyses	17	Report other analyses done—eg analyses of subgroups and interactions, and sensitivity analyses	Not applicable (protocol)
Discussion			
Key results	18	Summarise key results with reference to study objectives	Not applicable (protocol)

Limitations	19	Discuss limitations of the study, taking into account sources of potential bias or imprecision. Discuss both direction and magnitude of any potential bias	10
Interpretation	20	Give a cautious overall interpretation of results considering objectives, limitations, multiplicity of analyses, results from similar studies, and other relevant evidence	9,10
Generalisability	21	Discuss the generalisability (external validity) of the study results	10
Other information			
Funding	22	Give the source of funding and the role of the funders for the present study and, if applicable, for the original study on which the present article is based	10

*Give information separately for cases and controls in case-control studies and, if applicable, for exposed and unexposed groups in cohort and cross-sectional studies.
Note: An Explanation and Elaboration article discusses each checklist item and gives methodological background and published examples of transparent reporting. The STROBE checklist is best used in conjunction with this article (freely available on the Web sites of PLoS Medicine at http://www.plosmedicine.org/, Annals of Internal Medicine at http://www.annals.org/, and Epidemiology at http://www.epidem.com/). Information on the STROBE Initiative is available at www.strobe-statement.org.