

Bayesian factor analysis in primary care: a systematic review protocol

Hao Zhang (✉ hao.zhang5@mail.mcgill.ca)

McGill University <https://orcid.org/0000-0002-1144-1093>

Tibor Schuster

McGill University

Protocol

Keywords: Bayesian, factor analysis, primary care, family medicine

Posted Date: March 12th, 2020

DOI: <https://doi.org/10.21203/rs.3.rs-16959/v1>

License: © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Version of Record: A version of this preprint was published on January 8th, 2021. See the published version at <https://doi.org/10.1186/s13643-020-01565-6>.

Abstract

Background: Bayesian factor analysis allows for efficient use of preliminary data and information that corresponds to the increasing needs of questionnaire construct validation in primary care research. This systematic review will summarise evidence on the current use of Bayesian factor analysis in primary care.

Methods: We will adopt a comprehensive search strategy to identify relevant literature (research studies in primary care) indexed in PubMed, Medline, Scopus, EMBASE, CINAHL and Cochrane Library. The search strategy will include terms (and synonyms) for Bayesian statistics, factor analysis and primary care. We will conduct forward and backward searches manually on the references of articles that meet the inclusion and exclusion criteria to further identify eligible studies. Multiple reviewers will conduct data extraction independently. The analyses will include descriptive synthesis summarizing features about the use and reporting of the respective Bayesian factor analysis approach.

Discussion: This systematic review will provide a bird view of the current use of Bayesian factor analysis in primary care and provide recommendations for its proper future use in primary care and beyond.

Systematic review registration : PROSPERO registration number: CRD42018114978 **Keywords:** Bayesian, factor analysis, primary care, family medicine

Background

In the past few decades, there has been a proliferation of primary care research studies in many countries and with publications in the field increasing by about 75% and major primary care research databases approximately tripling the amount of information stored between 2004 and 2013 in just the UK (1, 2). Despite this noticeable growth of available health information in various health domains, ensuring sufficient quality, i.e., validity and reliability, of relevant data remains a major challenge (3, 4). In the context of primary care practice and research, efficient validation of questionnaire instruments is critical in order to respond timely to emerging needs for effective clinical assessments or for addressing urgent research questions and, more general, a learning healthcare system that builds upon evidence-based medicine.

Factor analysis examines the strength of correlation of each individual questionnaire item with respect to the latent (unobservable) domain or construct (i.e. the “factor”) and is widely used for questionnaire construct validation in education, psychological, social, and healthcare research. Bayesian methods allow for incorporation of stakeholder knowledge and opinions during the instrument development, enabling inclusion of health professionals and patients as partners as it is advocated in patient-centered medicine and patient-oriented research, increasing practical value and applicability of the instrument (5). The empirical validation of measurement properties of an instrument through factor analysis typically involves a relatively large sample of completed questionnaires that can be prohibitively expensive and time consuming to obtain. The feasibility of these necessary pilot inquiries therefore strongly depends on

practicalities surrounding recruitment. The suggested minimum sample sizes for conventional confirmatory factor analysis in the literature range between approximately 100 and 300 responses (6). An additional limiting element is that individuals who participate in pilot validation studies can typically not be involved in subsequent phases of the questionnaire instrument validation. This is critical as primary care research often applies in community settings and may be aiming at addressing the needs of populations that are comparatively low in numbers and difficult to identify and recruit.

Advanced Bayesian methods offer promising solutions to this impasse as the incorporation of prior information can increase efficiency compared to conventional methods, arriving at posterior estimates with meaningful precision (6-9). Nevertheless, when screening the literature, there is an underutilization of Bayesian factor analysis in primary care and the application and reporting of Bayesian factor analysis seem to vary largely. To our best knowledge, no systematic review has been undertaken to quantify and qualify the use of Bayesian factor analysis in primary care. We therefore aim to provide the first comprehensive systematic review on this matter.

The objectives of this systematic review are: (1) assess the quality of the implementation and reporting of Bayesian factor analysis in primary care practice and research settings; (2) identify and consolidate the existing Bayesian factor analysis approaches in primary care practice and research; and (3) conduct a synthesis of used approaches for prior elicitation and Bayesian inference, including different estimation procedures and software routines.

Methods/design

Inclusion Criteria

To be included in the systematic review, articles must report on Bayesian methods and factor analysis in the context of primary care practice or research. For the purpose of this literature review, the definition of primary care follows that of the American Academy of Family Physicians as being “comprehensive”, “first contact” and “continuing” meanwhile it covers “any undiagnosed sign, symptom, or health concern” (10). Family medicine is characterized as “an academic and scientific discipline” and “a clinical specialty” “orientated to primary care” (11). We are looking at the aspects of primary care as in family medicine, epidemiology, health services and policy research. We will use a high sensitivity search strategy to identify all potentially relevant records to family medicine and primary care (12, 13).

Types of studies

We will include quantitative and empirical research studies, methodological studies using Bayesian factor analysis, review articles, conference abstracts and thesis / dissertation documents. Research studies using similar model structures such as structure equation model and latent variable model, as well as item response theory, factor loading and item domain correlation with Bayesian methods in primary care context will be included. The full text paper of the included studies must be available. There is no language limitation for the articles to be included.

Exclusion Criteria

We will exclude editorials, commentaries, book reviews, hypotheses, critical appraisals, reflections, surveys, case reports or studies using other frequentist statistical methods (such as hypothesis tests, confidence intervals, and p-values) or with no information on Bayesian methods. Studies that include some of the key words but use them under different connotations or references are excluded. Examples of ineligible use of words include “primary studies”, “prior to”, “human epidermal growth factor” and “genetic factor”.

Bayesian methods used in other types of analyses, such as Bayes rule, Bayes / Bayesian factor studies, variational Bayes, Bayesian Information Criterion/Criteria, Bayesian random effects models, Bayesian/Bayes network, belief network and Bayes(ian) model or probabilistic directed acyclic graphical model are excluded. Instead of factor analysis, studies using hierarchical Poisson models with latent variables, Gaussian process and risk factor analysis are excluded. Studies not in family medicine or primary care but use related words are excluded. Examples include: “a family of methods” and “exponential family”.

If no information is given in the title or abstract about any of the three criteria, i.e. no indication about whether the study is Bayesian, using factor analysis or in primary care, we include those studies at the initial stage of screening. When not sure, especially the term “factor analysis” is mentioned but not specified whether it is Bayesian or not Bayesian, the article is kept for the next round of full-text review.

Search methods for identification of studies

A comprehensive search strategy is adopted to identify potential studies, indexed in PubMed and translated to other databases. The search strategy will include terms (and synonyms) for Bayesian, factor analysis and primary care. The search strategy will be developed with a specialized librarian and conducted by at least two reviewers independently. We will combine searches of electronic databases with hand searches of reference lists. The computer-based searches will combine medical subject headings (MeSH terms) and free text (or full text) searches related to Bayesian factor analysis in primary care and family medicine. We will systematically search for studies using Bayesian factor analysis reporting on the (i) data characteristics and (ii) use of Bayesian factor analysis from inception to date.

Electronic searches and time frame

We will identify articles from PubMed, Medline, Embase, Cochrane Library, CINAHL, and Scopus. We consider all relevant articles published before January 1st, 2020.

Searching other resources

Google Scholar will be manually scanned for the first 100 records for supplementary information. Reference lists and the future citation of the retrieved articles will be manually searched with two

additional rounds. Fig. 1 shows the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) flow diagram for the identification of studies (14).

Data collection and analysis

Selection of studies

Titles and abstracts of studies are sequentially screened using the search strategy by at least two independent reviewers using the software Rayyan applying the inclusion and exclusion criteria (15). Studies that represent a 'best fit' will be included. Deadline is given for the review to be done in a limited timeframe. The full text of articles that meet the inclusion criteria will be retrieved and examined independently by the reviewers. Any disagreement between the reviewers about the eligibility of specific studies will be discussed and additional reviewer will be involved if necessary, until consensus is reached. For studies with multiple publication records, the most comprehensive or up-to-date record will be used.

Data extraction and management

Data extraction are independently conducted by at least two reviewers. All records will be coded and categorized under the predefined themes (codebook) from the CIHR (Canadian Institute of Health Research) grants and rewards guide (16). Despite existed guidelines / recommendations on reporting of general Bayesian methods, confirmatory factor analysis and questionnaire development independently and separately, no single comprehensive recommendation was found on the reporting of Bayesian confirmatory factor analysis (17-19). Where applicable, the following data will be extracted: the types of journal, the publication dates, the geographical locations, the sample sizes, the number of items/questions used for the Bayesian factor analysis, the number of factors (domains/constructs), the reported item-domain correlations, regression parameters and/or factor loadings (parameters of structural equation models), the use of prior information and assumed prior distributions, and their primary care settings. A standardized predesigned data collection form will be used for data extraction. We will follow the screening criteria below:

1. Did the authors use either a Bayesian confirmatory factor analysis or Bayesian exploratory factor analysis or Bayesian latent variable model or Bayesian structure equation model?
2. If they used one of those, what was the parameter of interest they were aiming to estimate using a Bayesian approach: item-to-domain correlation, factor loading or latent model regression parameter? In other words, for which parameter did they impose a prior distribution?
3. How did they inform their prior distribution of the respective parameter? What was the prevalence of studies that employed non-informative priors?
4. If they mention the term "factor loading", did they explain it and if, how did they interpret it i.e. as item-to-domain correlation or as model parameter (latent variable coefficient)?
5. Did they report factor loadings (results) that exceeded [-1, 1] interval?

6. Were credible intervals (or confidence intervals) reported for factor loading / item-to-domain correlation / model parameter (regression coefficients)?
7. What software / libraries are used, are software codes or original data available? (reproducibility)

Assessment of quality of implementation and reporting of Bayesian methods

Risk of bias in individual studies is not applicable to and will not be assessed in our study since the goal is to summarize the use of Bayesian methods, i.e. there is no single effect parameter that is of primary interest. The data collected and pooled across studies are information on the presence or absence of specific criteria in the design, conduct and reporting of Bayesian factor analysis. We will assess the quality of implementation and reporting of Bayesian methods for each eligible study rated on a scale of very low, low, moderate and high on the following aspects: reporting about methodology, Bayesian model, estimated parameters, factor loading, informed prior, and basic information about the publication. The assessment of quality will be presented in tables in the final publication of the systematic review. Given that no available critical tools yet exist to appraise the use of Bayesian factor analysis, we will develop a recommended procedure / tool for the use and reporting of Bayesian factor analysis.

Strategy for data descriptions / synthesis

We will provide a descriptive-analytical synthesis of the findings from the included studies with graphs and tables detailing the use of Bayesian factor analysis based on a common analytical framework on authors, years of publication, estimates, the number of publications as changing over time, geographical locations, the study populations, the aims of the study, data types, key information about the data (sample sizes, number of questions in a questionnaire, number of domains/factors), the type of Bayesian method used, different estimation procedures and software routines (e.g. analytical solutions vs. sampling-based solutions) and important results. We will compare, summarize and report the results using content analysis based on the research themes in CIHR guidebook to examine the use of Bayesian factor analysis (16). We will chart the data, synthesizing and interpreting the data.

Statistical Synthesis

We anticipate conducting meta-analyses to summarize relative frequencies of Bayesian factor analysis approaches being used as well as general and specific reporting issues.

Results

Description of studies

The current use of Bayesian factor analysis will be summarized through descriptive statistics and content analysis. A guideline will be developed for methodological and practical procedures for better reporting

and interpreting Bayesian factor analysis results to assist researchers and clinicians relying on such findings in the context of primary care.

Discussion

This systematic review will provide a detailed summary of how Bayesian factor analysis methods are applied in primary care practice and research settings. It will enable identification of challenges in use / application and reporting of Bayesian factor analysis studies within the context of primary care and will help to improve practice through identifying needs for improving Bayesian factor analysis applications and reporting. No one single agreed definition of the research domain of primary care and family medicine yet exists, which might affect the search results.

Abbreviations

CINAHL Cumulative Index to Nursing & Allied Health Literature

EMBASE Excerpta Medica database

UK United Kingdom

PRISMA Preferred Reporting Items for Systematic Reviews and Meta-Analyses

Declarations

Ethics approval and consent to participate

Not applicable

Consent for publication

Not applicable

Availability of data and materials

Not applicable

Competing interests

The authors declare no competing interests

Funding

Hao Zhang was funded by *Fonds de la recherche en santé du Québec (FRQS)* and the *Quebec Strategy for Patient-Oriented Research Support for People and Patient-Oriented Research and Trials (SPOR-SUPPORT) Unit*. Dr. Schuster was supported through funds obtained from a *Canadian Institutes of Health Research Canada Research Chair (Tier II)* award. The funders had no direct impact in developing the protocol.

Authors' contributions

HZ and TS initiated and design the study. HZ drafted the manuscript. All authors read and approved for the final manuscript.

Acknowledgements

The work is funded by the Fonds de la recherche en santé du Québec and the Quebec-SPOR SUPPORT Unit at the Department of Family Medicine at McGill University.

Author's information

¹Department of Family Medicine, McGill University,

5858 Chemin de la Côte-des-Neiges, Suite 300, Montréal, QC H3S 1Z1, Canada

References

1. Chaudhry Z, Mannan F, Gibson-White A, Syed U, Ahmed S, Kousoulis A, et al. Outputs and growth of primary care databases in the United Kingdom: bibliometric analysis. *Journal of innovation in health informatics*. 2017;24(3):284-90.
2. Hajjar F, Saint-Lary O, Cadwallader J-S, Chauvin P, Boutet A, Steinecker M, et al. Development of Primary Care Research in North America, Europe, and Australia From 1974 to 2017. *The Annals of Family Medicine*. 2019;17(1):49-51.
3. Hierholzer Jr WJ. Health care data, the epidemiologist's sand: comments on the quantity and quality of data. *The American journal of medicine*. 1991;91(3):S21-S6.
4. Sanders J, Powers B, Grossmann C. *Digital Data Improvement Priorities for Continuous Learning in Health and Health Care: Workshop Summary*: National Academies Press; 2013.
5. Sacristán JA. Patient-centered medicine and patient-oriented research: improving health outcomes for individual patients. *BMC medical informatics and decision making*. 2013;13(1):6.
6. MacCallum RC, Widaman KF, Zhang S, Hong S. Sample size in factor analysis. *Psychological methods*. 1999;4(1):84.

7. Lee S-M, Abbott P, Johantgen M. Logistic regression and Bayesian networks to study outcomes using large data sets. *Nursing Research*. 2005;54(2):133-8.
8. Canadian Institute for Health Information. *A Performance Measurement Framework for the Canadian Health System (Updated November 2013)*: Canadian Institute for Health Information; 2013.
9. Garrard L, Price LR, Bott MJ, Gajewski BJ. A novel method for expediting the development of patient-reported outcome measures and an evaluation across several populations. *Applied psychological measurement*. 2016;40(7):455-68.
10. American Academy of Family Physicians. *Primary care*. 2019.
11. WONCA Europe. *The European definition of general practice/family medicine 2019*.
12. Pols DH, Bramer WM, Bindels PJ, van de Laar FA, Bohnen AM. Development and validation of search filters to identify articles on family medicine in online medical databases. *The Annals of Family Medicine*. 2015;13(4):364-6.
13. Gill PJ, Roberts NW, Wang KY, Heneghan C. Development of a search filter for identifying studies completed in primary care. *Family practice*. 2014;31(6):739-45.
14. Moher D, Liberati A, Tetzlaff J, Altman DG. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. *Annals of internal medicine*. 2009;151(4):264-9.
15. Ouzzani M, Hammady H, Fedorowicz Z, Elmagarmid A. Rayyan—a web and mobile app for systematic reviews. *Systematic reviews*. 2016;5(1):210.
16. Canadian Institutes of Health Research. *The Four Themes of CIHR Funded Health Research*. 2018.
17. Floyd FJ, Widaman KF. Factor analysis in the development and refinement of clinical assessment instruments. *Psychological assessment*. 1995;7(3):286.
18. Jackson DL, Gillaspay Jr JA, Purc-Stephenson R. Reporting practices in confirmatory factor analysis: An overview and some recommendations. *Psychological methods*. 2009;14(1):6.
19. Spiegelhalter DJ, Myles JP, Jones DR, Abrams KR. Bayesian methods in health technology assessment: a review. *Health Technology Assessment (Winchester, England)*. 2000;4(38):1-130.

Figures

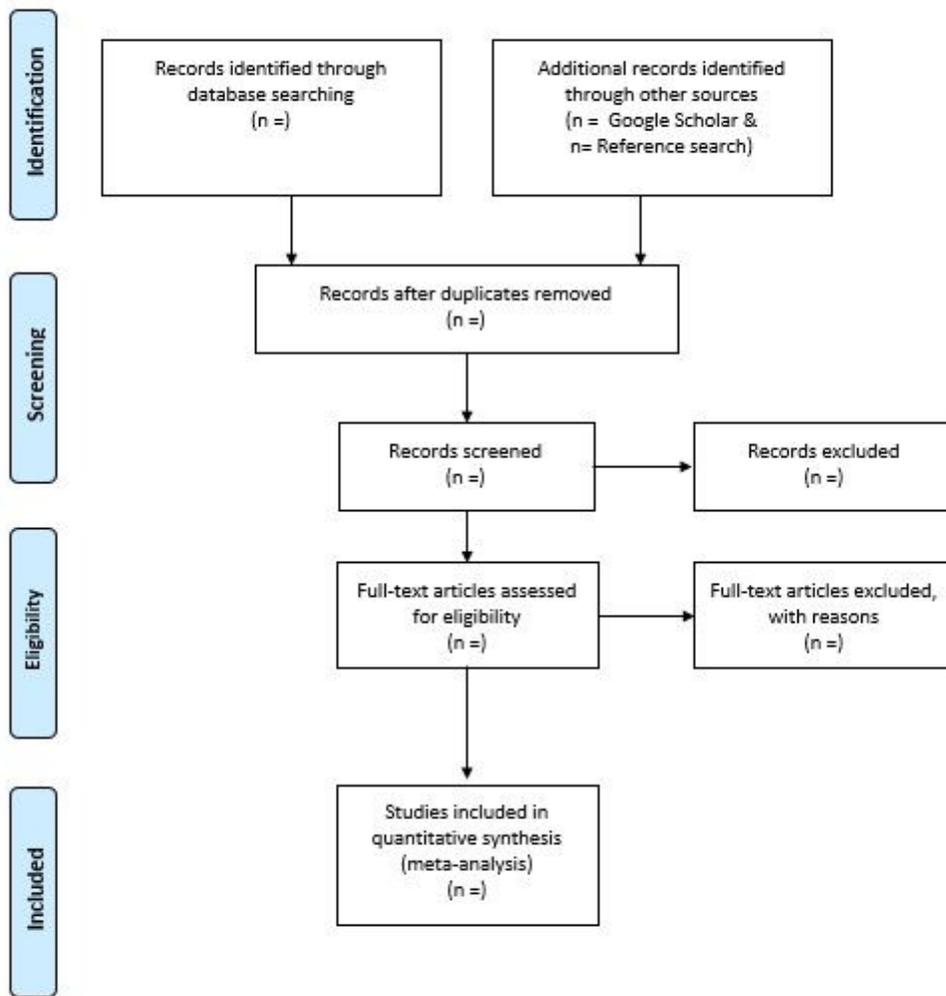


Figure 1

PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) flow diagram for the identification of studies

Supplementary Files

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

- [PRISMAPchecklist.docx](#)
- [Additionalfile1.docx](#)