PEER REVIEW HISTORY

BMJ Open publishes all reviews undertaken for accepted manuscripts. Reviewers are asked to complete a checklist review form and are provided with free text boxes to elaborate on their assessment. These free text comments are reproduced below.

ARTICLE DETAILS

Title (Provisional)

Developing a professional competency framework for general practitioners in tertiary hospitals in China: A modified Delphi study

Authors

Wei, Yun; An, Yanhua; Cao, Qiumei; Feng, Wei; Wang, Dawei; Zhu, Dan

VERSION 1 - REVIEW

Reviewer	**1**
Name	**Andreou, Vasiliki**
Affiliation	**KU Leuven, Department of Public Health and Primary Care**
Date	**02-Jan-2024**
COI	**None**

I would like to thank the authors fort his well-written manuscript. I really enjoyed reading the manuscript.

Abstract: In general, the abstract efficiently summarizes the article. Nevertheless, as a reader, I would like to read in the results more about the consensus ratings and which competencies deemed important by the experts, rather than the coefficients.

Introduction:

The introduction clearly frames the problem within the Chinese context. However, it is not clear to me what exactly this study aims to address. Is the competency framework necessary to address shortcomings in patient care or is it necessary to evaluate GPs? If the latter applies, how did the authors manage to tackle the context differences between tertiary and primary care? I wonder whether this information would be appropriate to be described under Methods section as study setting. In addition, it is not clear to me what exactly the role of the GPs is within tertiary hospitals. Are they first point of care within the hospital? Do they have tasks related to patients follow-up and continue of care? Also, I personally miss the connection with literature about competency-based education and continuing professional development. How is this study situated within competency-based education for GPs? How do the authors define competencies?

Methods:

L27, p6 How did the authors define agreement?

L48, P6: What was the necessary academic background that experts were required to have?

L50-51, p6: Delphi literature proposes to keep the group of experts as homogeneous as possible to avoid misconceptions about the research topic. Can the authors explain why they opted for a heterogeneous panel?

L58-59,p6: How were the participants invited to participate and by whom?

L9-11,p7: I would like to read here more information about the secondary and tertiary indicators and the structure of the framework (relation among the indicators).

L22-23; p7: Were the criteria for choosing the competencies evidence-based or were they chosen because of their relevance to the study context? Also, under questionnaire preparation, the authors use the term “indicators” for the first time. Are the indicators another term for competencies? I would propose to use coherent terminology throughout the manuscript to avoid confusion.

L41-42,p7: Why did the authors choose for a 9-point Likert scale?

L46-47,p7: How was the degree of experts’ familiarity assessed? Was the familiarity not considered as an inclusion criterium?

L48-50,p7: Were the participants explicitly asked to leave comments after each round?

Also, how was the questionnaire administered and validated? Which information was communicated to the participants in each round? Were there any instructions given for filling in the questionnaire? All this information can help reproducibility of the manuscript.

L5-7, p8: How do the authors define importance and feasibility? Could the authors include which exact questions the participants were asked?

L19, p8: I assume that the authors mean “agreement”, as they described in the design, when they write consensus. I would propose to use the term consensus throughout the manuscript since the Delphi methodology is a consensus based methodology. Also, how was consensus achieved?

L25-26,p8: Why were competencies removed when consensus was not achieved and not modified?

L47-52,p8: The information about consensus should be described under the design. That would clarify a lot of questions throughout the methods section.

L4-22,p9: This paragraph is not clear for the reader. The Delphi methodology is an opinion-based methodology, therefore, I do not see the point to try to standardize it. Unavoidably, there is a degree of bias in this methodology. Additionally, how were the judgement-making ability and familiarity measured? Furthermore, Kendall's W is not typically used directly

within the Delphi methodology. The Delphi method is a structured communication technique that aims to achieve a consensus among a group of experts on a particular issue. It involves a series of rounds in which experts provide feedback and revise their opinions based on the feedback received from the group. The goal is to converge towards a group consensus. While Kendall's W is a measure of agreement for ranked data, it does not directly capture the iterative and consensus-building process of the Delphi method. In the Delphi methodology, the focus is on reaching a convergence of expert opinions rather than assessing the agreement in their initial rankings. Could the authors explain why they took this methodological choice? Also, it seems that a third criterium was added in the questionnaire, namely sensitivity. Could the authors explain what do they mean with sensitivity?

L23-37, p9: How were qualitative comments analysed?

L27;p9: there is a typo “characteristics of participants”.

Results:

L58, p9: How do the authors define seniority? What is exactly a senior grade title?

Discussion

L9-48,p12: In my opinion, this paragraph does not contribute a lot of information to the manuscript. Literature has repeatedly risen concerns and addresses issues regarding the reliability of the Delphi methodology. However, these issues and concerns should inform methodological choices during the design of the study.

L35-47,p13: Can the authors explain how the teaching and research aspect of their framework is different from the CanMEDS role of Scholar?

L4-28,p14: Can the authors explain how competency domain prevention is different from the CanMEDS role of Health Advocate?

What about the importance of collaborating with other team members to ensure continuity of care? This aspect seems to be missing in the current framework. Could the authors explain what was the reason about it?

Author’s contributions: L8, p16: Can the authors clarify how these specific authors participated in the study? What do the authors mean by participating in study? Participation in the study as participant has implications for conflict of interest.

Table 3: Competency 1.4.1 has a typo “be empathic (listener)”.

Reviewer	**2**
Name	**Armson, Heather**
Affiliation	**University of Calgary**

Date 04-Jan-2024

COI Assistant Dean Office of Continuing Medical Education and Professional Development, Cumming School of Medicine, University of CalgaryExecutive Director, The Foundation for Medical Practice Education, McMaster University

---

Introduction: a clearer understanding of the proposed role of GPs in tertiary hospital settings would be helpful: are all patients seen at this level admitted to hospital or are there also outpatient clinics in this setting? Are these physicians functioning as 'hospitalists'? Is all GP residency training occurring at the tertiary care setting including for those physicians who plan to practice in primary or secondary settings? What is a standardized residency training program as discussed on page 6 line 27? Is the expectation of a PhD related to the research expectations at this level? How will the needed competencies differ from those developed for GPs after standardized residency training? (Page7 line 50)

Methods

A delphi model is appropriate assuming the choice of participants is sufficient. The authors identify some limitations related to the participants including stage of practice and location. However, I am also curious as to gender given that almost all participants are female. Is there a reason for this skew- are most GPs female? Why is the professional title important- is it a surrogate for years in practice?

Were there descriptions of the indicators provided to participants? For example, what does data processing mean?

What differentiates primary, secondary and tertiary indicators?

Statistical analysis:

I am unsure what is meant by 'the degrees of experts' activeness' (pg 10 line 60) as a measure of credibility and scientific basis of the results. How was the judgment making ability and familiarity with surveyed indications (Cr & Cs) identified as there is not description in the analysis outlining data collected to address these measures?

Results

There was great retention of all participants. Was there any difference in rating between GPs across the < 5years, 5-10 years & over 10 years in practice groups? Were the results consistent between those physicians who had on the job training compared to those with standardized or other residency training? It maybe that the sample size is too small to address these elements.

Discussion

Appropriate comparison to other competency frameworks.

Conclusion

Consistent with results.

English Language-some minor changes in language are required for clarity throughout the document: For example in Supplementary material 1.1.1 & 1.1.2 "Be with" at the beginning of the indicators does not make sense. 1.4.1 'empathy' instead of empathetic.

Reviewer	**3**
Name	**Beyene, Bereket**
Affiliation	**Hawassa University College of Medicine and Health Sciences, School of Nursing**
Date	**01-May-2024**
COI	**no competing interest**

Comment to author

1. The author should design appropriate method for studying the research question?

2. I also recommend qualitative approach from the government body to study the research

3. The study result should be defined clearly based on the objective

---

Reviewer	**4**
Name	**Zhang , Anqing**
Affiliation	**Children's National Medical Center**
Date	**05-Aug-2024**
COI	**N/A**

Review on "Developing a professional competency framework for general practitioners in tertiary hospitals in China: A modified Delphi study"

Manuscript ID: bmjopen-2023-082736

Reviewer: Anqing Zhang, PhD

Reviewer's research areas: Biostatistics, Clinical trial design and analysis, Oncology studies, Dose-response study.

Summary:

This manuscript investigated the medical service in general practice (GP) departments of tertiary hospitals in China and developed a professional competency framework for GPs in tertiary hospitals in China using a modified Delphi study method.

This paper had excellent response rates for both rounds when conducted the Delphi survey.

Research questions, study method/design, and conclusion are clearly stated and addressed. Below are some comments for the authors:

Comments for authors:

1) Considering the relatively small sample size, should report median and interquartile range (IQR) instead of mean and standard deviation for the numerical variables, for example, age of experts.
2) The minimum number of samples for Delphi study requires to be at least 30 to provide rigor for statistical analysis. Suggest the author states this in the limitation section.
3) In the questionnaire preparation stage, how many indicators in total were searched in the literature review? The manuscript showed there are 4 primary indicators, 12 secondary indicators and 48 tertiary indicators were selected, however, we have no idea how many in total were reviewed and where were they from. Recommend to report the numbers of original indicators from each source that were initially reviewed to reflect the rationale of the selections.
4) recommend to report the IQR along with median to reflect the variation of the median scores for Table 3.

Reviewer	**5**
Name	**Shang, Zhida**
Affiliation	**McGill University**
Date	**10-Oct-2024**
COI	**NONE**

The selection criteria for the expert panel are excellent, as it has experts from a variety of roles related to general practice in tertiary hospitals. This allows for diverse and informed perspectives.

A high participation rate (19 out of 20 invited experts) is a strong point, as it reduces bias and enhances the credibility of the consensus.

Although Delphis with 2 rounds are acceptable, the best practice for Delphi nowadays are 3 rounds in the medical literature. The study could have discussed further the reasoning behind stopping at two rounds.

VERSION 1 - AUTHOR RESPONSE

Responses to the reviewer 1

Point 1: Abstract: In general, the abstract efficiently summarizes the article. Nevertheless, as a reader, I would like to read in the results more about the consensus ratings and which

competencies deemed important by the experts, rather than the coefficients. **Response:** The results in the **Abstract** have been revised. Please see "**Abstract**, line 10-27, page 2".

Point 2: Introduction: The introduction clearly frames the problem within the Chinese context. However, it is not clear to me what exactly this study aims to address. Is the competency framework necessary to address shortcomings in patient care or is it necessary to evaluate GPs? If the latter applies, how did the authors manage to tackle the context differences between tertiary and primary care? I wonder whether this information would be appropriate to be described under Methods section as study setting. In addition, it is not clear to me what exactly the role of the GPs is within tertiary hospitals. Are they first point of care within the hospital? Do they have tasks related to patients follow-up and continue of care? Also, I personally miss the connection with literature about competency-based education and continuing professional development. How is this study situated within competency-based education for GPs? How do the authors define competencies?

Response: Generally, general practice provides "person-centred, continuing, comprehensive and coordinated whole person healthcare to individuals and families in their communities with common and frequently-occurring diseases".[19] A multi-method study exploring the work content of GPs in primary care in Beijing indicated that GP-patient consultation with common and frequently-occurring diseases is the major part of GP work. In addition, GPs also undertake work like chronic disease management and follow-up, health file management, family doctor contract services, teaching student, etc.[20] In addition to outpatient medical services, GPs in tertiary hospitals also deliver inpatient diagnostic and therapeutic services, which is different from primary healthcare providers that exclusively offer outpatient care. Typically, the epidemiology of multimorbidity among the outpatients and inpatients admitted to the general

practice departments of tertiary hospitals are complex. As reported in previous study, the prevalence of multimorbidity among inpatients in the general practice department of tertiary hospitals in China is extremely high, reaching 93.1%.[7] In tertiary hospitals, in addition to providing clinical diagnosis and treatment within both outpatient and inpatient departments, GPs are also required to engage in educational activities related to standardized resident training, conduct scientific research pertinent to career advancement and professional title promotion, as well as participate in disease prevention and management initiatives. These responsibilities have established heightened expectations for the competencies of GPs working in tertiary hospitals. Given the varied backgrounds of GPs in tertiary hospitals and the current focus of competency evaluations primarily on GPs in primary care, there is a notable absence of literature addressing competency evaluation for GPs in tertiary hospitals and the associated evaluation tools. Consequently, this study aims to establish a professional competency framework for GPs in tertiary hospitals, thereby providing a reference point for future assessments of GP competencies. Please see "**Introduction**, line 13-30, page 7, line 1-8, page 8".

GPs in tertiary hospitals appreciate the importance of research, actively engaging in and applying it within their practice to ensure they remain competent to deliver high-quality, evidence-based care that supports positive patient and population health outcomes. The scientific research capabilities of GPs are also closely linked to continuing medical education and continuing professional development in China.[38] Please see "**Discussion-*Comparison to previous competency frameworks***, line 26-30, page 17".

Competency-based education is particularly well-suited for the course design of GP training

programs and the assessment of competency after training. This study underscores the significance of teaching ability for GPs in tertiary hospitals, thereby rendering this competency model less applicable.

Professional competency in medicine was defined as "the habitual and judicious use of communication, knowledge, technical skills, clinical reasoning, emotions, values, and reflection in daily practice for the benefit of the individual and community being served" by Epstein and Hundert in JAMA.[12] This study does not propose a novel definition of post-competency.

Point 3: Methods: L27, p6 How did the authors define agreement?

Response: The process concludes upon reaching a consensus regarding the topics under discussion. The Delphi study lacked definitive consensus criteria[27]. In this study, consensus was established based on two selection parameters: a median score exceeding seven on a nine-point scale and at least 70% of panel ratings falling within the top tertile (7–9) for both importance and feasibility[28]. Please see "**Methods-** ***Design***, line 19-23, page 8".

Point 4: L48, P6: What was the necessary academic background that experts were required to have?

Response: (i) expert authority, which means the academic background related to general practice in tertiary hospitals, including roles in leading or participating in research, seminars, and academic conferences related to the establishment, positioning, and development of general practice departments. Please see "**Methods-** ***Participants***, line 4-8, page 9".

Point 5: L50-51, p6: Delphi literature proposes to keep the group of experts as homogeneous as possible to avoid misconceptions about the research topic. Can the authors explain why they opted for a heterogeneous panel?

Response: Although Delphi literature proposes to keep the group of experts as homogeneous as possible to avoid misconceptions about the research topic, experts from a wide range of sources in this study offer a more comprehensive perspective. Furthermore, despite the experts' diverse professional backgrounds, they share a commonality in possessing a comprehensive understanding of the functional orientation of general practice department and the responsibilities of GPs in tertiary hospitals.

Point 6: L58-59,p6: How were the participants invited to participate and by whom?

Response: Finally, 20 eligible experts were invited by Professor Cao via E-mail and 19 experts agreed to participate in this study. Please see "**Methods-** ***Participants***, line 16-17, page 9".

Point 7: L9-11,p7: I would like to read here more information about the secondary and tertiary indicators and the structure of the framework (relation among the indicators).

Response: A new supplementary material about "Preliminary professional competency framework for GPs in tertiary hospitals in China" is provided. Please see "**Supplementary material 2.** Preliminary professional competency framework for GPs in tertiary hospitals in China".

Point 8: L22-23; p7: Were the criteria for choosing the competencies evidence-based or were

they chosen because of their relevance to the study context? Also, under questionnaire preparation, the authors use the term "indicators" for the first time. Are the indicators another term for competencies? I would propose to use coherent terminology throughout the manuscript to avoid confusion.

Response: Potential competency indicators were extracted and screened by 2 reviewers (YW and YHA) according to following criteria: (a) the indicators were applicable to measure the competency of GPs in tertiary hospitals; (b) the indicators were relevant to requirements of GPs' work in tertiary hospitals in China; (c) the indicators were relevant to development of GPs in tertiary hospitals. When there were doubts about whether an indicator should be retained, the research team would discuss together to make a decision. Please see "**Methods-*Participants***, line 4-10, page 10".

To enhance clarity, we have opted to utilize "competency indicators" to articulate the three-tiered metrics within the competency model.

Point 9: L41-42, p7: Why did the authors choose for a 9-point Likert scale?

Response: Although 3-point [1#], 5-point [2#] and 7-point [3#] scales have also been used, the 9-point Likert scale is frequently employed for ratings [4#,5#]. Therefore, 9-point Likert scale been chosen in this study.

Point 10: L46-47,p7: How was the degree of experts' familiarity assessed? Was the familiarity not considered as an inclusion criterium?

Response: The familiarity of experts indicates their comprehensive understanding of the meanings and contexts related to competency indicators, which are evaluated on a 1-5 Likert scale (1 = not familiar; 5 = very familiar). The familiarity was considered when in inclusion

criterium as "(ii) a wide range of sources, including management personnel in general practice departments in tertiary hospitals, GPs in tertiary hospitals, government administrators, or scientific researchers in the field of general practice, who possessed a comprehensive understanding of the responsibilities of GPs in tertiary hospitals".

However, in light of the fact that multiple experts articulate skepticism concerning the part the scientific soundness and rationality of the Delphi method, and most previous articles reporting Delphi results failed to address this aspect, we decide to remove this section in the revised manuscript.

Point 11: L48-50,p7: Were the participants explicitly asked to leave comments after each round? Also, how was the questionnaire administered and validated? Which information was communicated to the participants in each round? Were there any instructions given for filling in the questionnaire? All this information can help reproducibility of the manuscript.
Response: Spaces were left for experts to make comments on these existing competency indicators or recommend new competency indicators which they considered should be included in. Please see "**Methods-** ***Questionnaire preparation***, line 26-28, page 10".

The first-round questionnaire was sent to experts by e-mail, along with materials about the research background, the aim of the study, the demographic information collection form, instructions of scoring criteria, and descriptions of the indicators. In the first-round questionnaire, experts were asked to rate the importance and feasibility of each competency indicator using the 1-9 Likert scale, give their comments on the existing indicators, and recommend new competency indicators which they considered should be included in. Please see "**Methods-** ***Delphi survey***, line 1-7, page 11".

The competency indicators confirmed in the first round of Delphi survey were formulated into the second-round questionnaire, which was sent to the same experts with the first-round survey by e-mail, accompanied by a graph-based report detailing the results from the first round. Importance and feasibility of each level of competency indicators were rated using the same 1-9 Likert scale as in the first round. In this round of survey, participants were also given a chance to suggest additional competency indicators, argue for or against proposed competency indicators, and comment on competency indicators wording and comprehension. Please see "**Methods-** ***Delphi survey***, line 18-26, page 11".

Point 12: L5-7, p8: How do the authors define importance and feasibility? Could the authors include which exact questions the participants were asked? **Response:** Importance pertains to the significance of the indicator in reflecting the competencies of GPs and feasibility pertains to the accessibility of information concerning evaluation outcomes during the actual evaluation process, which were both rated on a 1-9 Likert scale (1 = not important/feasible at all; 9 = very important/feasible)[27,28]. Please see "**Methods-** ***Questionnaire preparation***, line 20-24, page 10".

Definition about importance and feasibility was provided in "instructions of scoring criteria" section of the two-round expert consultation questionnaire.

Point 13: L19, p8: I assume that the authors mean "agreement", as they described in the design, when they write consensus. I would propose to use the term consensus throughout the manuscript since the Delphi methodology is a consensus based methodology. Also, how was

consensus achieved?

Response: The process concludes upon reaching a consensus regarding the topics under discussion. The Delphi study lacked definitive consensus criteria[27]. In this study, consensus was established based on two selection parameters: a median score exceeding seven on a nine-point scale and at least 70% of panel ratings falling within the top tertile (7–9) for both importance and feasibility[28]. Please see "**Methods-** ***Design***, line 19-23, page 8".

Point 14: L25-26,p8: Why were competencies removed when consensus was not achieved and not modified?

Response: The competency indicators that fail to achieve consensus suggest a low level of both importance and feasibility, rendering them unsuitable for the competency evaluation of GPs in tertiary hospitals.

Point 15: L47-52,p8: The information about consensus should be described under the design. That would clarify a lot of questions throughout the methods section.
Response: The information about consensus has been described under the design in revised manuscript.

Point 16: L4-22,p9: This paragraph is not clear for the reader. The Delphi methodology is an opinion-based methodology, therefore, I do not see the point to try to standardize it. Unavoidably, there is a degree of bias in this methodology. Additionally, how were the judgement-making ability and familiarity measured? Furthermore, Kendall's W is not typically used directly within the Delphi methodology. The Delphi method is a structured communication technique that aims to achieve a consensus among a group of experts on a

particular issue. It involves a series of rounds in which experts provide feedback and revise their opinions based on the feedback received from the group. The goal is to converge towards a group consensus. While Kendall's W is a measure of agreement for ranked data, it does not directly capture the iterative and consensus-building process of the Delphi method. In the Delphi methodology, the focus is on reaching a convergence of expert opinions rather than assessing the agreement in their initial rankings. Could the authors explain why they took this methodological choice? Also, it seems that a third criterium was added in the questionnaire, namely sensitivity. Could the authors explain what do they mean with sensitivity?
Response: In light of the fact that multiple experts articulate skepticism concerning the part the scientific soundness and rationality of the Delphi method, and most previous articles reporting Delphi results failed to address this aspect, we decide to remove this section in the revised manuscript.

Point 17: L23-37, p9: How were qualitative comments analysed?
Response: All qualitative feedback from experts will be systematically extracted and categorized into distinct groups, encompassing revisions to the descriptions of indicators, proposed deletions of certain indicators, and suggestions for new indicators to be added. The occurrence frequency of identical suggestions will be recorded. Please see "**Methods-*Statistical analysis***, line 23-27, page 12".

Point 18: L27;p9: there is a typo "characteristics of participants".
Response: It has been corrected in the revised version. Please see "**Methods-*Statistical analysis***, line 18, page 12".

Point 19: Results: L58, p9: How do the authors define seniority? What is exactly a senior grade title?

Response: In China, the professional titles for physicians are categorized into four distinct levels: junior grade (resident physician), intermediate grade (attending physician), deputy senior grade (deputy chief physician), and senior grade (chief physician). These classifications are determined by the healthcare professionals' work experience and research accomplishments. Please see the "note" part in "**Table 1**. Panel characteristics of the Delphi process".

Point 20: Discussion L9-48,p12: In my opinion, this paragraph does not contribute a lot of information to the manuscript. Literature has repeatedly risen concerns and addresses issues regarding the reliability of the Delphi methodology. However, these issues and concerns should inform methodological choices during the design of the study.
Response: This part has been removed in the revised manuscript.

Point 21: L35-47,p13: Can the authors explain how the teaching and research aspect of their framework is different from the CanMEDS role of Scholar?
Response: In comparison to the foreign competency models for GPs in America,[13] Australia,[16] and Europe,[35] the professional competency framework for GPs in tertiary hospitals in China imposes more stringent requirements regarding teaching. As clinical residential training bases, general practice department in tertiary hospitals need to undertake tasks about teaching and training, including taking the lead in formulating and implementing training plans, carrying out outpatient and ward teaching, cooperating with primary care institutions in teaching.[8] Although teaching ability is also emphasized in the CanMEDS role of Scholar,[15] teaching activities and competence requirements of GPs in tertiary hospitals in China are mainly focused on clinical

practice, thus facilitating the transition of residency trainees from theoretical knowledge to practical application. Furthermore, aside from a few trainees engaged in the general practice department in tertiary hospitals, the majority of trainees pursue their careers within primary care institutions after residency training. Consequently, the ability of joint teaching with primary care is crucial not only for aiding students in mastering clinical skills in hospitals but also for considering the case characteristics and diagnostic approaches relevant to primary healthcare. Please see "**Discussion-** ***Comparison to previous competency frameworks***, line 4-22, page 17".

GPs in tertiary hospitals appreciate the importance of research, actively engaging in and applying it within their practice to ensure they remain competent to deliver high-quality, evidence-based care that supports positive patient and population health outcomes. The scientific research capabilities of GPs are also closely linked to continuing medical education and continuing professional development in China.[38] Similar to the CanMEDS role of Scholar,[15] competencies related to research design, implementation, and the translation of research findings have been underscored in the competency framework for GPs in tertiary hospitals, which are not adequately represented in competency models from America,[13] Australia,[16] and Europe.[35] Please see "**Discussion-** ***Comparison to previous competency frameworks***, line 26-30, page 17, line 1-5, page 18".

Point 22: L4-28,p14: Can the authors explain how competency domain prevention is different

from the CanMEDS role of Health Advocate?

Response: Another important indicator of professional competency framework for GPs in tertiary hospitals in China was prevention. The provision of effective preventive care aims to reduce preventable morbidity and mortality, enhance quality of life and decrease an individual's need generally for medical services.[39] Since the mid-1990s professional bodies have argued that prevention should be a constituent element of normal professional practice of GPs and nurses and that prevention and health promotion should be an integral part of general practice.[40] GPs can positively influence their patient's lifestyle choices, and encourage and equip them to take a greater interest in, and greater responsibility for, their own health.[41] Same as the family medicine milestone project in America,[13] the role of health advocate as outlined in CanMEDS from Canada,[15] and the competency profile of Australian general practitioner at the point of fellowshi,[16] disease prevention, encompassing screening and health risks management, constitutes a critical component of competency evaluation of GPs in tertiary hospitals across China. Please see "**Discussion-** ***Comparison to previous competency frameworks***, line 9-22, page 18".

Point 23: What about the importance of collaborating with other team members to ensure continuity of care? This aspect seems to be missing in the current framework. Could the authors explain what was the reason about it?

Response: Collaborating with other team members is important. GPs in tertiary hospitals are required to engage in collaborative efforts with various hospital departments as well as primary healthcare institutions. Although this study did not include indicators for team collaboration, GPs communication ability with colleagues from different departments in hospital and primary

care institutions were emphasized. in addition, the significance of joint educational initiatives with primary care was also highlighted.

Point 24: Author's contributions: L8, p16: Can the authors clarify how these specific authors participated in the study? What do the authors mean by participating in study? Participation in the study as participant has implications for conflict of interest. **Response:** YW and QMC designed the study. YHA, WY, and QMC participated in Delphi questionnaire preparation and data collection. All authors collaboratively developed and refined the Delphi questionnaire. The database was established and inputted by WF, DZ, and DWW. QMC would check and correct it If there was any difference or error. YW wrote the manuscript and revised it according to the reviewer's comments. QMC reviewed and revised the manuscript. All authors read and approved the final manuscript. Please see "**Authors' contributions**, line 22-27, page 20".

Point 25: Table 3: Competency 1.4.1 has a typo "be empathic (listener)". **Response:** It has been corrected in the revised version. Please see "Supplementary material 3. Results of the Delphi process for tertiary competency indicators".

Responses to the reviewer 2

Point 1: Introduction: a clearer understanding of the proposed role of GPs in tertiary hospital settings would be helpful: are all patients seen at this level admitted to hospital or are there also outpatient clinics in this setting? Are these physicians functioning as 'hospitalists'? Is all GP residency training occurring at the tertiary care setting including for those physicians who plan to practice in primary or secondary settings? What is a standardized residency training program as discussed on page 6 line 27? Is the expectation of a PhD related to the research expectations at this level? How will the needed competencies differ from those developed for GPs after

standardized residency training? (Page7 line 50)

Response: Generally, general practice provides "person-centred, continuing, comprehensive and coordinated whole person healthcare to individuals and families in their communities with common and frequently-occurring diseases".[19] A multi-method study exploring the work content of GPs in primary care in Beijing indicated that GP-patient consultation with common and frequently-occurring diseases is the major part of GP work. In addition, GPs also undertake work like chronic disease management and follow-up, health file management, family doctor contract services, teaching student, etc.[20] In addition to outpatient medical services, GPs in tertiary hospitals also deliver inpatient diagnostic and therapeutic services, which is different from primary healthcare providers that exclusively offer outpatient care. Typically, the epidemiology of multimorbidity among the outpatients and inpatients admitted to the general practice departments of tertiary hospitals are complex. As reported in previous study, the prevalence of multimorbidity among inpatients in the general practice department of tertiary hospitals in China is extremely high, reaching 93.1%.[7] In tertiary hospitals, in addition to providing clinical diagnosis and treatment within both outpatient and inpatient departments, GPs are also required to engage in educational activities related to standardized resident training, conduct scientific research pertinent to career advancement and professional title promotion, as well as participate in disease prevention and management initiatives. These responsibilities have established heightened expectations for the competencies of GPs working in tertiary hospitals. Please see "**Introduction**, line 13-30, page 7, line 1-2, page 8".

The standardized residency training represents the primary pathway for GP training. Upon successful completion of the residency program, trainees will be eligible to register as GPs and pursue careers in community health service institutions (CHSIs) or within the general practice departments of hospitals. The standardized residency training comprises two distinct phases: (1) 30 months dedicated to hospital-based clinical rotations, and (2) 6 months focused on CHSI-

based training. [9,10] Please see "**Introduction**, line 24-29, page 6".

Point 2: Methods: A delphi model is appropriate assuming the choice of participants is sufficient. The authors identify some limitations related to the participants including stage of practice and location. However, I am also curious as to gender given that almost all participants are female. Is there a reason for this skew- are most GPs female?

Response: Another limitation of this study is that, despite our efforts to recruit male participants, the majority of participated experts were female. This imbalance can be attributed to the predominance of female practitioners in clinical medicine and medical education in China, particularly within the fields of internal medicine, gynecology, pediatrics, and general practice. Please see "**Discussion-*Strengths and limitations***, line 26-30, page 19".

Point 3: Why is the professional title important- is it a surrogate for years in practice?

Response: In China, the professional titles for physicians are categorized into four distinct levels: junior grade (resident physician), intermediate grade (attending physician), deputy senior grade (deputy chief physician), and senior grade (chief physician). These classifications are determined by the healthcare professionals' work experience and research accomplishments. Please see the "note" part in "**Table 1**. Panel characteristics of the Delphi process".

Point 4: Were there descriptions of the indicators provided to participants? For example, what does data processing mean?

Response: The first-round questionnaire was sent to experts by e-mail, along with materials

about the research background, the aim of the study, the demographic information collection form, instructions of scoring criteria, and descriptions of the indicators. Please see "**Methods - *Delphi survey*** line 1-4, page 11".

Point 5: What differentiates primary, secondary and tertiary indicators? **Response:** Four primary competency indicators (medical services, teaching, research, and prevention) were determined based on the basic functions of the general practice departments in tertiary hospitals.[8] The secondary indicators pertain to the dimensions of competency evaluation, while the tertiary indicators refer to the specific content associated with competency assessment.

Point 6: Statistical analysis: I am unsure what is meant by 'the degrees of experts' activeness' (pg 10 line 60) as a measure of credibility and scientific basis of the results. How was the judgment making ability and familiarity with surveyed indications (Cr & Cs) identified as there is not description in the analysis outlining data collected to address these measures? **Response:** In light of the fact that multiple experts articulate skepticism concerning the part about the scientific soundness and rationality of the Delphi method, and most previous articles reporting Delphi results failed to address this aspect, we decide to remove this section in the revised manuscript.

Point 7: Results: There was great retention of all participants. Was there any difference in rating between GPs across the <5years, 5-10 years & over 10 years in practice groups? Were the results consistent between those physicians who had on the job training compared to those with standardized or other residency training? It maybe that the sample size is too small to address these elements.

Response: As you mentioned, the study was not subjected to further stratification due to its limited sample size.

Point 8: Discussion: Appropriate comparison to other competency frameworks. **Response:** The professional competency framework for GPs in tertiary hospitals across China has been discussed compared with the foreign competency models for GPs in America, Australia, Canada, and Europe, as well as competency assessment tools in China. Please see "**Discussion - *Comparison to previous competency frameworks,*** line 11-30, page 16, line 1-30, page 17, line 1-30, page 18, and line 1-9, page 19".

Point 9: English Language-some minor changes in language are required for clarity throughout the document: For example in Supplementary material 1.1.1 & 1.1.2 "Be with" at the beginning of the indicators does not make sense. 1.4.1 'empathy' instead of empathetic.

Response: Grammar and typos have been corrected in the revised manuscript.

Responses to the reviewer 3

Point 1: The author should design appropriate method for studying the research question? **Response:** A modified Delphi method was adopted in the study, which was the most widely used method for selecting quality indicators in healthcare.[21,22]

Point 2: I also recommend qualitative approach from the government body to study the research.

Response: (ii) a wide range of sources, including management personnel in general practice departments in tertiary hospitals, GPs in tertiary hospitals, government administrators, or scientific researchers in the field of general practice, who possessed a comprehensive understanding of the responsibilities of GPs in tertiary hospitals; Please see "**Methods –**

Participants, line 8-11, page 9".

Point 3: The study result should be defined clearly based on the objective.

Response: The result of the Delphi consultation was reported and the scientific soundness and rationality of the Delphi method was removed in the revised manuscript.

Responses to the reviewer 4

Point 1: Considering the relatively small sample size, should report median and interquartile range (IQR) instead of mean and standard deviation for the numerical variables, for example, age of experts.

Response: Mean and standard deviation for the numerical variables have been changed to median and interquartile range (IQR) in the revised manuscript.

Point 2: The minimum number of samples for Delphi study requires to be at least 30 to provide rigor for statistical analysis. Suggest the author states this in the limitation section.

Response: As reported in previous study that a sample of about fifteen has been suggested [6#,7#] and larger panels have also been used.

Point 3: In the questionnaire preparation stage, how many indicators in total were searched in the literature review? The manuscript showed there are 4 primary indicators, 12 secondary indicators and 48 tertiary indicators were selected, however, we have no idea how many in total were reviewed and where were they from. Recommend to report the numbers of original indicators from each source that were initially reviewed to reflect the rationale of the selections.

Response: Four primary competency indicators (medical services, teaching, research, and prevention) were determined based on the basic functions of the general practice departments in tertiary hospitals.[8] A preliminary list of secondary and tertiary competency indicators was constructed by literature review. Literature was searched in PubMed and three Chinese

databases (China National Knowledge Infrastructure, Wanfang Data, VIP Chinese Periodical Services) with terms commonly used to describe GP (e.g., general practitioner, family physician, family doctor), tertiary hospital (e.g., tertiary hospital, general hospital, hospital), competency (e.g., competency, competence, ability). Furthermore, policy documents related to GP in tertiary hospitals across China were also reviewed to extract competency indicators. Finally, a total of 31 published research papers describing domestic and foreign GPs' competencies were identified form literature review, which included 5 published competency models from international general practice organizations. In addition, 3 published policy documents about general practitioner system in China were also reviewed (references of these papers and policies were shown at supplementary material 1).

Potential competency indicators were extracted and screened by 2 reviewers (YW and QMC) according to following criteria: (a) the indicators were applicable to measure the competency of GPs in tertiary hospitals; (b) the indicators were relevant to requirements of GPs' work in tertiary hospitals in China; (c) the indicators were relevant to development of GPs in tertiary hospitals. When there were doubts about whether an indicator should be retained, the research team would discuss together to make a decision. There were 74 competency indicators identified by the screening process. After deleting duplicate competency indicators, integrating the indicators with similar dimensions, and classifying them into three hierarchical levels based on their connotations, a preliminary professional competency framework for GPs in tertiary hospitals in China was conducted including 4 primary indicators, 14 secondary indicators and 48 tertiary indicators (Supplementary material 2). Please see "**Methods – *Questionnaire preparation,*** line 19-30, page 9, line 1-16, page 10".

Point 4: recommend to report the IQR along with median to reflect the variation of the median scores for Table 3.

Response: Given the configuration of the table and the ineffectiveness of the IQR in achieving consensus, it has not been incorporated into Table 3 (Supplementary material 3. Results of the Delphi process for tertiary competency indicators) at this time.

Responses to the reviewer 5

Point 1: The selection criteria for the expert panel are excellent, as it has experts from a variety of roles related to general practice in tertiary hospitals. This allows for diverse and informed perspectives.

A high participation rate (19 out of 20 invited experts) is a strong point, as it reduces bias and enhances the credibility of the consensus.

Response: Thank you for your positive comments.

Point 2: Although Delphis with 2 rounds are acceptable, the best practice for Delphi nowadays are 3 rounds in the medical literature. The study could have discussed further the reasoning behind stopping at two rounds.

Response: There is no restriction on the number of rounds that can be conducted,[23] but two or three rounds are most common in previous studies. [24-26] The process concludes upon reaching a consensus regarding the topics under discussion in this study, which was established based on two selection parameters: a median score exceeding seven on a nine-point scale and at least 70% of panel ratings falling within the top tertile (7–9) for both importance and feasibility[28].

References

7. [dataset] Zhou Z, Shi M, Liu M, et al. Data from: Multimorbidity in Hospitalized Patients Admitted to General Practice Departments and Its Implications for the General Practice Healthcare System: A Four-Year Longitudinal Study in China. Front Public Health. 2021 Dec 20;9:760792. doi: 10.3389/fpubh.2021.760792.
8. Department of Science and Technology Education. Guideline for the Establishment of General Practice in Standardized Training Base for Resident Doctors (General Hospital) (Trial) [in Chinese]. 2018. http://www.nhc.gov.cn/qjjys/s3593/201809/951a65647c41459b858ccf1c26fc1acb.shtml (accessed May 28, 2023)
9. Shao S, Wu T, Guo A, et al. The training contents, problems and needs of doctors in urban community health service institutions in China. BMC Fam Pract. 2018;19(1):182. doi: 10.1186/s12875-018-0867-6.
10. Chinese Medical Doctor Association. The content and standards for standardized residency training

(revised in 2019)- general practice training rules [in Chinese]. 2019. http:// www. cmda. net/u/ cms/ www/ 201911/12161 04487 4h. pdf. Accessed 13 Oct 2024.

12. Epstein RM, Hundert EM. Defining and assessing professional competence. JAMA. 2002;287(2):226–235. doi: 10.1001/jama.287.2.226.

13. The Accreditation Council for Graduate Medical Education and The American Board of Family Medicine. The Family Medicine Milestone Project. 2015. http://www.acgme.org/Portals/0/PDFs/Milestones/FamilyMedicineMilestones.pdf. Accessed February 17, 2021(accessed May 28, 2023)

15. The College of Family Physicians of Canada. CanMEDS-FM 2017: A competency framework for family physicians across the continuum. 2017. https://www.cfpc.ca/canmedsfm/. (accessed May 28, 2023)

16. The Royal Australian College of General Practitioners. Competency profile of the Australian general practitioner at the point of fellowship. 2015.

19. The Royal Australian College of General Practitioners. What is General Practice? http://www.racgp.org.au/whatisgeneralpractice. (accessed May 28, 2023)

20. Wei Y, Wang F, Pan Z, et al. Work Content of General Practitioners in Beijing, China: A Multi-method Study. Front Public Health. 2022 Apr 29;10:870224. doi: 10.3389/fpubh.2022.870224.

23. Veziari Y, Kumar S, Leach M. The development of a survey instrument to measure the barriers to the conduct and application of research in complementary and alternative medicine: a Delphi study. BMC Complement Altern Med. 2018;18(1):335. doi: 10.1186/s12906-018-2352-0.

24. Keeney S, Hasson F, Mckenna H. Consulting the oracle: ten lessons from using the Delphi technique in nursing research. J Adv Nurs. 2006;53(2):205–12. doi: 10.1111/j.1365-2648.2006.03716.x.

25. Jin G, Wei Y, Liu Y, et al. Development of type 2 diabetes mellitus quality indicators in general practice by a modified elphi method in Beijing, China. BMC Fam Pract. 2020;21(1):146. doi: 10.1186/s12875-020-01215-9.

26. Gagliardi D, Rondinone BM, Mirabile M, et al. The perspective of European researchers of national occupational safety and health institutes for contributing to a European research agenda: a modified Delphi study. BMJ Open. 2017;7(6):e015336. doi: 10.1136/bmjopen-2016-015336.

27. Diamond IR, Grant RC, Feldman BM, et al. Defining consensus: a systematic review recommends methodologic criteria for reporting of Delphi studies. J Clin Epidemiol. 2014;67(4):401-9. doi: 10.1016/j.jclinepi.2013.12.002.

28. [dataset] Bisson JI, Tavakoly B, Witteveen AB, et al. Data from: TENTS guidelines: Development of post-disaster psychosocial care guidelines through a Delphi process. British Journal of Psychiatry. 2010,196: 69-74. doi: 10.1192/bjp.bp.109.066266.

35. Wonca Europe. The European definition of general practice/family medicine [EB/OL].2011. https://www.woncaeurope.org/file/0dce3f9f-e60f-4416-b257-4698590bd0c9/WONCA_European_Definitions_2_v7%20(1).pdf. (accessed May 28, 2023)

38. Sherman L, Kuang M, Yang DD, et al. An Overview of Continuing Medical Education/Continuing

Professional Development Systems in China: A Mixed Methods Assessment. J CME. 2024 Jun 9;13(1):2363855. doi: 10.1080/28338073.2024.2363855.

39. The Guide to Clinical Preventive Services 2014: Recommendations of the U.S. Preventive Services Task Force. Rockville (MD): Agency for Healthcare Research and Quality (US); 2014 May. Report No.: 14-05158.
40. Sarah Gear. The Complete MRCGP Study Guide, 4th Edition, Healthy People: promoting health and preventing disease. CRC Press. 2012.
41. Royal Australian College of General Practitioners: Putting prevention into practice. Guidelines for the implementation of prevention in the general practice setting. 2nd ed. Melbourne: RACGP; 2006.

[1#]. Cassar Flores A, Marshall S, Cordina M. Use of the Delphi technique to determine safety features to be included in a neonatal and paediatric prescription chart. Int J Clin Pharm. 2014;36(6):1179–89. doi: 10.1007/s11096-014-0014-y.

[2#]. McDermott JH, Caiola SM, Kuhn KF, Stritter FT, Beza J. A delphi survey to identify the components of a community pharmacy clerkship. Am J Pharm Educ. 1995;59(4):334–41.

[3#]. Cantrill JA, Sibbald B, Buetow S. Indicators of the appropriateness of long term prescribing in general practice in the United Kingdom: consensus development, face and content validity, feasibility and reliability. Qual Health Care. 1998;7:130–5. doi: 10.1136/qshc.7.3.130.

[4#]. Campbell SM, Cantrill JA, Roberts R. Prescribing indicators for UK general practice: Delphi consultation study. BMJ.2000;321:1–5. doi: 10.1136/bmj.321.7258.425.

[5#]. Aljamal MS, Ashcroft D, Tully MP. Development of indicators to assess the quality of medicines reconciliation at hospital admission: an e-Delphi study. Int J Pharm Pract. 2016 May;24(3):209-16. doi: 10.1111/ijpp.12234.

[6#] Linstone HA, Turoff M. The Delphi survey: method techniques and applications. Reading: Addison-Wesley; 1975.

[7#] McMillan SS, King M, Tully MP. How to use the nominal group and Delphi techniques. Int J Clin Pharm. 2016 Jun;38(3):655-62. doi: 10.1007/s11096-016-0257-x. Epub 2016 Feb 5.

Note: # These studies referred only in this 'Point-by-point response to reviewers', which are not in the "Changes marked" version.

VERSION 2 - REVIEW

Reviewer	**1**
Name	**Andreou, Vasiliki**
Affiliation	**KU Leuven, Department of Public Health and Primary Care**
Date	**22-Dec-2024**

COI

Thank you for revising the manuscritp and addressing all the points.

Reviewer	**4**
Name	**Zhang , Anqing**
Affiliation	**Children's National Medical Center**
Date	**11-Dec-2024**
COI	

The authors have revised the manuscript per my comments, no further comments.