PEER REVIEW HISTORY

BMJ Open publishes all reviews undertaken for accepted manuscripts. Reviewers are asked to complete a checklist review form (**http://bmjopen.bmj.com/site/about/resources/checklist.pdf**) and are provided with free text boxes to elaborate on their assessment. These free text comments are reproduced below.

ARTICLE DETAILS

TITLE (PROVISIONAL)	Comparative survey among pediatricians, nurses, and health information technicians in ethics implementation knowledge and attitude of social experiments based on medical artificial intelligence at children's hospitals in Shanghai: a cross-sectional study
AUTHORS	Wang, Yingwen; Fu, Weijia; Gu, Ying; Fang, Weihan; Zhang, Yuejie; Jin, Cheng; Yin, Jie; Wang, Weibing; Xu, Hong; Ge, Xiaoling; Ye, Chengjie; Tang, Liangfeng; Fang, Jinwu; Wang, Daoyang; Su, Ling; Wang, Jiayu; Zhang, Xiaobo; Feng, Rui

VERSION 1 – REVIEW

REVIEWER	John Federick C Yap Holy Angel University, School of Nursing and Allied Medical Sciences
REVIEW RETURNED	25-Jan-2023

GENERAL COMMENTS	The title does not reflect the objective of the paper. Based on the current title, it is descriptive in nature. The comparison is not identified up until the statistical analysis section. What is the basis for comparing the 3 groups? Limit the use of decimal points to the tenths value for uniformity. In the measures section, was the tool used adapted or developed by the authors? In the results section, under the professional titles, the total of the percentage is 99.8 Replace with a dash on the results with 0 value. Given the results of table 2 and 3. What was the basis for the merging of the values to just 3 responses for the Chi-square analysis? In the conclusion, there was no mention of the Chi-square implications. In the results section on page 3, the term awareness was used instead of attitude.

REVIEWER	Christian C. Rose Stanford University
REVIEW RETURNED	07-Mar-2023

GENERAL COMMENTS	Thank you for submitting your research. Reading through this study, it appears to be a cross-sectional assessment of pediatric physician, nurse, and technologist opinions regarding the ethical implementation of medical artificial intelligence. However, it is difficult to clearly interpret the objective of this research, as well as why specific questions were chosen, likely secondary to the quality of writing. There are many typos throughout the abstract, as well as many places where punctuation is missing, or incorrect. Some

	sentences appear to start out of the blue without relation to the preceding information. I do, generally, however, believe that this work does a good job of illuminating that this particular cohort feels unprepared to address these ethical issues, and could utilize more training on the topic. That being said, it was not clear to me how representational this group is of pediatrics.

REVIEWER	Simone Grassi Università Cattolica del Sacro Cuore, Department of Health Surveillance and Bioethics
REVIEW RETURNED	12-May-2023

GENERAL COMMENTS	Verifying and enhancing compliance with norms and guidelines it is a core component of the "safety culture" from both an ethical and legal point of view, therefore I think that the objectives of the research group are of scientific interest. However, I think introduction, methodology, results and discussion deserve critical revision. Regarding introduction, there are more legal and ethical issues related to processing data of a minor and use artificial intelligence in pediatric context than those expressed by the authors (many of them are summarized in: 10.3389/fmed.2021.821756). In particular, I suggest to focus on and extensively discuss the international principle of the "best interest of the minor" and its trade-off with the risks entailed by current AI products (e.g., do the authors think that data immortality is a relevant issue for minors?). Regarding methods, I would be interested to have a clear indication of all the exact exclusion criteria. Moreover, it should be specified what the authors exactly mean for junior, intermediate, senior etc and in the results section it should be expressed (both in the text and in a table) the relationship between "seniority" and results of interest. In the discussion I would give more space to the critical discussion of all the results, also in the light of a critical comparison with previous evidence

REVIEWER	Alya Arabi United Arab Emirates University, Biochemistry and Molecular Biology
REVIEW RETURNED	25-May-2023

GENERAL COMMENTS	This is a study about the knowledge of healthcare worker on the ethics related to AI application in pediatrics. Below are the comments provided to the authors: Please define acronyms at first occurrence in the abstract (unless the Journal format does not allow it). English: It can be improved in many spots, there are typos and sentence structures that can be improved. For example, “Regarding the knowledge on ethical management of MAI application, it was asked by 10 questions” The sentence structure is difficult to follow. The sample size is small. Why is the Yamane’s formula in particular chosen to decide on the sample size? “Participants were excluded if their answering time was less than 150 seconds which was based on a pilot survey, or their answers were illogical (such as same answers to all items, unreasonable answers to birthday, etc)” what does etc refer to here? What was the pilot survey? how was the threshold 150 seconds reached? “Prior the study, we performed factor analysis for construct validity, calculated the interrater agreement (IRR), item-level content validity index(I-CVI), scale-level CVI (S-

	CVI) for content validity, and Cronbach alpha coefficient for internal consistency." need references for each calculated variable. "The questionnaire was pre-tested in a convenience sample of 6 healthcare workers. Based on their feedback, minor changes were made to enhance clarity and appropriateness of the questions." A lot more details need to be provided here. Why only 6 participants? Have they participated in the questionnaire after this pre-test? What kind of feedback did they give? How did the changes made enhance the clarity and appropriateness of the questions? What are the questions included in the questionnaire? Based on what were they prepared/decide on? Are the statements included in e.g. Table 2 and Table 3 the questions themselves? If yes, they are really vague and lack precision. It would have been beneficial to include some questions that are related to specific ethics points and/or laws/legislations. What are the unfamiliar, uncertain, and familiar in Table 4? They need to be clearly defined in the text. Also, why giving the option of 5 different answers if they were to be merged into three categories, e.g. disagree, neutral, and agree? Why not, for example, give three possible answers from the beginning? "It showed that the overall knowledge -level of all participants was low, which was consistent with the results of Zheng". What is Zheng's study about? what is in particular the consistency? Why are the patients excluded from the study while they are the most affected by knowing the ethics of the AI used in their access to healthcare? The discussion is very superficial and lacks depth. It also lacks more comparisons with other studies (there was only one comparison with another study, and it was not clear). The strengths and limitations are limited to limitations with only two general ones being listed (bias and small sample size). Overall the study is weak, nothing innovative to it, little informative as obvious from the conclusion.

VERSION 1 – AUTHOR RESPONSE

Reviewer: 1

Dr. John Federick C Yap, Holy Angel University

Comments to the Author:

The title does not reflect the objective of the paper. Based on the current title, it is descriptive in nature. The comparison is not identified up until the statistical analysis section.

RESPONSE : We appreciate the reviewer's kind suggestion. We have revised the title of the manuscript as advised: Comparative survey among pediatricians, nurses, and health information technicians in ethics implementation knowledge and attitude of social experiments based on medical artificial intelligence at children's hospitals in Shanghai: a cross-sectional study.

What is the basis for comparing the 3 groups?

RESPONSE : We appreciate the reviewer's question. Pediatricians, nurses, and health information technicians have more opportunities to be potential researchers in MAI social experiments. Their ethics implementation knowledge and attitudes are vital in mitigating ethical risks and then may

influence decision-making processes and pediatric patient care. We have added a brief explanation in the 'Introduction' section(See Page 5, line 104-106).

Limit the use of decimal points to the tenths value for uniformity.
RESPONSE： We appreciate the reviewer's kind suggestion. We have revised the data in our manuscript as advised.

In the measures section, was the tool used adapted or developed by the authors?
RESPONSE： We appreciate the reviewer's question. Yes, the questionnaire was developed by us. We have added details about how the questionnaire in supplementary materials.
.

In the results section, under the professional titles, the total of the percentage is 99.8
RESPONSE： We apologized for the mistakes. We have corrected the data of the percentage of intermediate for 24.4% (See Page 9, Table 1).

Replace with a dash on the results with 0 value.
RESPONSE： We appreciate the reviewer's kind suggestion. We have revised the data in our manuscript as advised.

Given the results of table 2 and 3. What was the basis for the merging of the values to just 3 responses for the Chi-square analysis?
RESPONSE： We appreciate the reviewer's kind suggestion and question. We have revised the table 2 and 3, with all responses to 3 grades.

In the conclusion, there was no mention of the Chi-square implications.
RESPONSE： We appreciate the reviewer's suggested question. We have totally rewritten the conclusion section(See Page 17-18, Line 288-294).

In the results section on page 3, the term awareness was used instead of attitude.
RESPONSE： We appreciate the reviewer's kind reminder. We have revised the term awareness for attitude.

Reviewer: 2
Dr. Christian C. Rose, Stanford University, Stanford University School of Medicine
Comments to the Author:
Thank you for submitting your research. Reading through this study, it appears to be a cross-sectional assessment of pediatric physician, nurse, and technologist opinions regarding the ethical implementation of medical artificial intelligence. However, it is difficult to clearly interpret the objective of this research, as well as why specific questions were chosen, likely secondary to the quality of writing. There are many typos throughout the abstract, as well as many places where punctuation is missing, or incorrect. Some sentences appear to start out of the blue without relation to the preceding information. I do, generally, however, believe that this work does a good job of illuminating that this particular cohort feels unprepared to address these ethical issues, and could utilize more training on the topic. That being said, it was not clear to me how representational this group is of pediatrics.
RESPONSE： We apologize for the grammatical and tone problems in our manuscript, we have rechecked the whole article once again and revised them based on your recommendations. In addition, we have invited a native English editor (Ms. Rui Wang) to polish and modify the manuscript

and acknowledged her assistance in the "Acknowledgement" section. We have also totally rewritten the 'Introduction', 'Method', 'Results', and 'Discussion' sections to clarify the study.

Reviewer: 3
Dr. Simone Grassi, Università Cattolica del Sacro Cuore
Comments to the Author:
Verifying and enhancing compliance with norms and guidelines it is a core component of the "safety culture" from both an ethical and legal point of view, therefore I think that the objectives of the research group are of scientific interest. However, I think introduction, methodology, results and discussion deserve critical revision. Regarding introduction, there are more legal and ethical issues related to processing data of a minor and use artificial intelligence in pediatric context than those expressed by the authors (many of them are summarized in: 10.3389/fmed.2021.821756). In particular, I suggest to focus on and extensively discuss the international principle of the "best interest of the minor" and its trade-off with the risks entailed by current AI products (e.g., do the authors think that data immortality is a relevant issue for minors?). Regarding methods, I would be interested to have a clear indication of all the exact exclusion criteria. Moreover, it should be specified what the authors exactly mean for junior, intermediate, senior etc and in the results section it should be expressed (both in the text and in a table) the relationship between "seniority" and results of interest. In the discussion I would give more space to the critical discussion of all the results, also in the light of a critical comparison with previous evidence
RESPONSE : We appreciate the reviewer's kind suggestion. We have totally rewritten the paper. In the 'Introduction' section, we have added more legal and ethical issues related to processing data of a minor and use artificial intelligence in pediatric context than in10.3389/ fmed. 2021. 821756 (see Page 5, Line 95-103). In the 'Method' section, we added a clear indication of all the exact exclusion criteria (see Page 6-7, Line 126-133). The meaning for junior, intermediate, senior etc and in the results section was expressed (see Page 9, Line 179-182). The 'Discussion' section was totally rewritten.

Reviewer: 4
Dr. Alya Arabi, United Arab Emirates University
Comments to the Author:
This is a study about the knowledge of healthcare worker on the ethics related to AI application in pediatrics. Below are the comments provided to the authors:
Please define acronyms at first occurrence in the abstract (unless the Journal format does not allow it).
RESPONSE : We appreciate the reviewer's kind suggestion. We have define 'MAI' at the first occurrence in the abstract (see Page 2, Line 24-25).

English: It can be improved in many spots, there are typos and sentence structures that can be improved. For example, "Regarding the knowledge on ethical management of MAI application, it was asked by 10 questions" The sentence structure is difficult to follow.
RESPONSE : We apologize for the grammatical and tone problems in our manuscript, we have rechecked the whole article once again and revised them based on your recommendations. In addition, we have invited a native English editor (Ms. Rui Wang) to polish and modify the manuscript and acknowledged her assistance in the "Acknowledgement" section.

The sample size is small. Why is the Yamane's formula in particular chosen to decide on the sample size?
RESPONSE : We appreciate the reviewer's question. We understand your concern regarding the small sample size in our study and your question about why Yamane's formula was specifically chosen to determine the sample size. Allow us to address this query and provide you with a satisfactory explanation.

Yamane's formula, also known as the Yamane's sample size formula, is a widely used method for determining sample sizes in certain research studies, particularly in the field of survey research. It is a simplistic approach that offers a quick and easy way to estimate the required sample size, given a large population.
In our specific study, we chose to employ Yamane's formula to calculate the sample size due to the following reasons:
Convenience: Yamane's formula offers a straightforward and practical approach to estimating the sample size, requiring minimal resources and time compared to more complex sampling methods.
Population Size: The formula is particularly useful when dealing with large populations, where it becomes challenging and often impractical to survey every individual. By using Yamane's formula, we can obtain a representative sample that can yield meaningful results.
Level of Precision: Yamane's formula takes into account the desired level of precision or acceptable margin of error. This allows us to strike a balance between obtaining a sample size that is statistically significant and feasible within the constraints of our study.
While Yamane's formula is a popular choice for certain research contexts, we acknowledge that it may not be suitable for all study designs or situations. Depending on the research objectives, population characteristics, and available resources, alternative methods for determining sample size, such as power analysis or stratified sampling, may be more appropriate.
In future studies, we will consider the limitations of Yamane's formula and carefully evaluate the suitability of alternative methods to ensure robust and reliable results.

"Participants were excluded if their answering time was less than 150 seconds which was based on a pilot survey, or their answers were illogical (such as same answers to all items, unreasonable answers to birthday, etc)" what does etc refer to here? What was the pilot survey? how was the threshold 150 seconds reached?
RESPONSE : Thank you for the reviewer's valuable question. Data from the participants were excluded in the final analysis, once the answering time recorded by the Application was less than 150 seconds. This criterion was based on our pilot survey. During the pilot survey, it was observed that participants required a minimum of 150 seconds to complete the questionnaire adequately. Additionally, participants who provided the same response to all items were also excluded from the analysis.

"Prior the study, we performed factor analysis for construct validity, calculated the interrater agreement
(IRR), item-level content validity index(I-CVI), scale-level CVI (S-CVI) for content validity, and Cronbach alpha coefficient for internal consistency." need references for each calculated variable.
RESPONSE : We appreciate the reviewer's kind suggestion. We have added references for each calculated variable (see Page 7, Line 147-149).

"The questionnaire was pre-tested in a convenience sample of 6 healthcare workers. Based on their feedback, minor changes were made to enhance clarity and appropriateness of the questions." A lot more details need to be provided here. Why only 6 participants? Have they participated in the questionnaire after this pre-test? What kind of feedback did they give? How did the changes made enhance the clarity and appropriateness of the questions?
RESPONSE : We appreciate the reviewer's question. We have added details about how the questionnaire was developed (See Supplementary Materials)

What are the questions included in the questionnaire? Based on what were they prepared/decide on? Are the statements included in e.g. Table 2 and Table 3 the questions themselves? If yes, they are really vague and lack precision. It would have been beneficial to include some questions that are related to specific ethics points and/or laws/legislations.

RESPONSE : We appreciate the reviewer's question. We have provided the questionnaire in Supplementary Materials.

What are the unfamiliar, uncertain, and familiar in Table 4? They need to be clearly defined in the text. Also, why giving the option of 5 different answers if they were to be merged into three categories, e.g. disagree, neutral, and agree? Why not, for example, give three possible answers from the beginning?
RESPONSE : We appreciate the reviewer's kind suggestion. We have revised the confused statements in Supplementary Materials and the'Resluts' section.

"It showed that the overall knowledge -level of all participants was low, which was consistent with the results of Zheng". What is Zheng's study about? what is in particular the consistency?
RESPONSE : We appreciate the reviewer's kind suggestion. We have clarified the statements in the 'Discussion' section (see Page 14, Line 228-229).

Why are the patients excluded from the study while they are the most affected by knowing the ethics of the AI used in their access to healthcare?
RESPONSE : We appreciate the reviewer's question. Pediatricians, nurses, and health information technicians have more opportunities to be potential researchers in MAI social experiments. Their ethics implementation knowledge and attitudes are vital in mitigating ethical risks and then may influence decision-making processes and pediatric patient care. We have clarified this in the 'Introduction' section (see Page 5, Line 104-107).

The discussion is very superficial and lacks depth. It also lacks more comparisons with other studies (there was only one comparison with another study, and it was not clear).
The strengths and limitations are limited to limitations with only two general ones being listed (bias and small sample size).
RESPONSE : We appreciate the reviewer's kind suggestion. We have totally rewritten the 'Discussion' section, and we really hope it can meet your requirements.

VERSION 2 – REVIEW

REVIEWER	John Federick C Rap Holy Angel University, School of Nursing and Allied Medical Sciences
REVIEW RETURNED	14-Jul-2023

GENERAL COMMENTS	I would like to clarify the information in Table 1. Specifically the number of pediatricians in relation to the education level. A medical degree is equivalent to a post-graduate degree. There are 137 pediatricians, but only 103 Master's degree and 15 Doctorate degree holders. Please clarify this discrepancy.

REVIEWER	Alya Arabi United Arab Emirates University, Biochemistry and Molecular Biology
REVIEW RETURNED	03-Aug-2023

GENERAL COMMENTS	Thank you for addressing many of the comments. However, unfortunately, as noted below, many of them are not clearly addressed. All acronyms need to be defined at first occurence. The authors commented about Yamane's formula, but did not justify the approprietness of its use in this study, which is a serious

	concern provided how small the sample size is. Still no suficient details are provided about the pilot study. Not including patients, who are at the centre of the importance of knowing the ethics, is a big caveat. Justifying why nurses, etc. are included does not address why patients were excluded. The details as to why 6 participants, and the associated questions are still not fully provided. It is not clear how the confusion of "unfamiliar, uncertain, and familiar, is addressed in the supplementary material. The discussion still remains superficial with very little comparisons with the literature. The limitations are also not transparent in highlighting e.g. that the small sample size is very small.

VERSION 2 – AUTHOR RESPONSE

Response to reviewer 2: Please comment on the representativeness of the included sample of pediatricians, nurses, and health information technicians.

RESPONSE： We apologize for any confusion about the representativeness of the included sample. Shanghai, serving as a model city where MAI has been applied in the social environment., has published local norms and regulations to govern MAI deployments. MAIs play a crucial role in assisting patient tirage, disease diagnosis, treatment, and follow-up within children's hospitals especially the two chosen tertiary ones with the most resources and support for AI. Pediatricians and nurses in the two hospitals are actively engaged as key healthcare providers or researchers in MAI social experiments. Additionally, in accordance with hospital policy, medical information technicians play a vital supporting role in ensuring the success of MAI social experiments. The primary aim of this study is to investigate the knowledge and attitudes of pediatric medical staff regarding ethics implementation in MAI social experiments, so we recruited pediatricians, pediatric nurses, and medical information technicians as the included sample.

Reviewer: 1

Dr. John Federick C Yap, Holy Angel University

Comments to the Author:

I would like to clarify the information in Table 1. Specifically the number of pediatricians in relation to the education level. A medical degree is equivalent to a post-graduate degree. There are 137 pediatricians, but only 103 Master's degree and 15 Doctorate degree holders. Please clarify this discrepancy.

RESPONSE： We apologize for any confusion about the information in Table 1. We appreciate the opportunity to clarify the medical education pathway in China for undergraduate, master's, and doctoral levels. In China, high school graduates undergo the Chinese College Entrance Exam, commonly referred to as the "gaokao," to become eligible for admission to medical colleges. The undergraduate program usually spans five years and covers a comprehensive curriculum in medical sciences, clinical training, and healthcare ethics. Upon successful completion, graduates are conferred with a bachelor's degree in medicine. Subsequently, after passing the medical licensing examination, and registering as a licensed doctor, individuals are legally permitted to practice medicine. Nevertheless, many students opt to further their education by enrolling in a master's program, which offers the opportunity to work in prestigious hospitals. This stage generally involves three years of specialized study and research in a specific medical field, such as surgery, internal medicine, or pediatrics. Graduates are awarded either a master's degree in medicine or a master's

degree in science in medicine. For those seeking the highest level of medical education, pursuing a doctoral degree is a viable option. This phase of education often extends over three to four years and involves in-depth research, the publication of scientific papers, and the completion of a dissertation. Notably, according to the Chinese educational system, someone who doesn't have a full-time post-graduate education experience can also be awarded a master or doctoral degree but without a diploma. Moreover, becoming a pediatrician is not always the primary choice for medical graduates in China, resulting in variations in educational levels among pediatricians.

In the survey using the questionnaire written in Chinese, we specifically inquired about the participants' academic qualifications, that is the types of diploma, and we have updated the terminology used to describe this information in Table 1.

Reviewer: 4

Dr. Alya Arabi, United Arab Emirates University

Comments to the Author:

Thank you for addressing many of the comments. However, unfortunately, as noted below, many of them are not clearly addressed.

All acronyms need to be defined at first occurence.

RESPONSE : We are so grateful for the reviewer's kind question. We have rechecked the whole manuscript once again and made sure that all acronyms are defined at first occurrence.

The authors commented about Yamane's formula, but did not justify the approprietness of its use in this study, which is a serious concern provided how small the sample size is.

Response: We appreciate the reviewer for bringing up this essential question. When we determined the sample size, we sought guidance from a statistician at the Clinical Trial Unit of our hospital, who assisted us in utilizing PASS 2021(test for one mean) for calculation. Setting alpha level at 0.05, power at 0.9, with the assumption of a 20% increase (or decrease) in the mean compared to the baseline, we arrived at a sample size of 227(as shown in the image below). Meanwhile, we also used the adjusted Yamane's formula to calculate the sample size, resulting in a figure of 226. We sincerely apologize for the manuscript's earlier incorrect figure of '266'.

Still no suficient details are provided about the pilot study.

RESPONSE : We apologize for the limited details provided before. We have updated the description in appendix A (see line 39-46, page 2).

Not including patients, who are at the centre of the importance of knowing the ethics, is a big caveat. Justifying why nurses, etc. are included does not address why patients were excluded.

RESPONSE: We value the reviewer's question and feedback. Initially, the primary aim of this study was to investigate the ethics implementation knowledge and attitudes of pediatric medical staff who serve as either medical healthcare providers or researchers in MAI social experiments. However, as pointed out by the reviewers, patients are the recipients of healthcare services facilitated by MAI, and they are at the center of the importance of knowing the ethics. Consequently, we plan to expand our research to include patients in future studies.

The details as to why 6 participants, and the associated questions are still not fully provided.

RESPONSE : We apologize for not meeting the requirements of the reviewer. First, we would like to clarify that we have revised the manuscript's incorrect figure of 6 participants to 8 participants in the last revision. That 8 participants were chosen was according to the cognitive debriefing guidelines provided by the Patient-Reported Outcomes Measurement Information System (PROMIS) Translation Director from our previous study asked for at least 7 participants, and the information we got from the 8 participants has already reached the data saturation. The details were provided in Appendix A (see line 21-32, page 1-2).

It is not clear how the confusion of "unfamiliar, uncertain, and familiar, is addressed in the supplementary material.

RESPONSE : We apologize for any confusion regarding the terms 'unfamiliar, uncertain, and familiar.' In the context of the knowledge dimension, respondents were asked to express their familiarity with various aspects, including the progress, ethical issues related to conducting MAI social experiments, and ethics governance according to norms and principles for such experiments. To clarify: 'Familiar' indicates the respondents possess knowledge of the subject. 'Unfamiliar' implies that respondents have limited knowledge. 'Uncertain' signifies that respondents are unsure whether they possess knowledge about the topic or not; they may have some level of doubt regarding their knowledge. We appreciate your understanding and hope this clarification helps.

The discussion still remains superficial with very little comparisons with the literature.

The limitations are also not transparent in highlighting e.g. that the small sample size is very small.

RESPONSE : We apologize for not meeting the requirements of the reviewer. Previous studies on implementing ethics in MAI social experiments have always paid more attention to regulating researchers, programmers, engineers, and data scientists in the stages of research, design, and development, but failed to notice that ethical issues in healthcare staff are equally important. We are the early research team dedicated to focusing on implementing ethics in MAI social experiments within the field of pediatrics, so we did have limited opportunities for direct comparisons with existing literature. However, we did make efforts to draw parallels with similar study findings (as indicated in line 247-252, page 13-14 in clear copy). Additionally, we have updated the limitation section to include considerations regarding sample size and representativeness of the sample.