# scientific reports

OPEN

# Interpretations of reproducibility crisis in medical education research: a qualitative study

Soleiman Ahmady[1,2], Noushin Kohan[3✉], Hadi Hamidi[4], Ahmad Vahednasiri[5] & Manijeh Hooshmandja[6]

The scientific community is currently grappling with the "reproducibility crisis," a phenomenon characterized by a significant loss of validity in research outcomes and a subsequent decline in public and scientific confidence. This crisis is evident in the high rate of studies, particularly within medicine and the humanities that fail replication tests and exhibit poor generalizability. This study aimed to specifically examine the issues of reproducibility in medical education research in Iran and propose effective solutions to address this challenge. This investigation employed a qualitative methodology, utilizing conventional content analysis. Data were collected through in-depth, semi-structured interviews with a carefully selected group of 24 medical science professors. Interviewing continued until data saturation was achieved, with each session lasting between 60 and 90 min. The analysis was systematically performed using the established Graneheim and Lundman's framework. The analysis led to the identification of three themes: factors contributing to the reproducibility crisis, its consequences, and potential solutions. These themes were further divided into eight categories. Contributing factors encompassed research methodology, various forms of bias, and contextual elements. The consequences were primarily reflected in two domains—scientific progress and decision-making in educational contexts. In contrast, proposed solutions centered on enhancing methodological rigor and reinforcing a culture of accountability and oversight in research practice. According to the findings, improving the quality and credibility of research in medical education requires a comprehensive strategy built upon three key pillars: advancing researchers' methodological competence, institutionalizing ethical and transparent research practices, and strengthening systems of evaluation and accountability. The effective implementation of initiatives such as the development of clear research guidelines, organization of training and capacity-building workshops, and promotion of robust inter-institutional collaboration can play a vital role in enhancing the quality, reliability, and trustworthiness of medical education research in Iran

**Keywords** Reproducibility crisis, Medical education, Research methodology, Scientific publications, Quality education

The scientific community has recently been faced with a concerning phenomenon termed the 'reproducibility crisis. At its core, this crisis signifies a substantial loss of validity in research outcomes, leading to diminished confidence in the generated findings. This issue is multifaceted, spanning various disciplines and challenging the fundamental principles of scientific inquiry, which rely on the ability of independent researchers to verify reported results[1].

The broad scope of this challenge is evident in estimates across various fields; for instance, it is widely acknowledged that only about 10–25% of biomedical research outcomes can be reliably reproduced, a statistic that has prompted considerable attention to reproducibility in biomedical research. These figures are employed to illustrate the general scope of the reproducibility crisis across research types (e.g., biomedical and social sciences) and are not strictly confined to any single methodology or discipline[2]. Scholars are increasingly

[1]School of Medical Education and Learning Technology, Shahid Beheshti University of Medical Sciences, Tehran, Iran. [2]Department of LIME, Research Affiliated Faculty, Karolinska Institute, Solna, Sweden. [3]Department of Medical Education and E-Learning in Medical Education, Smart University of Medical Sciences, Tehran, Iran. [4]Department of English Language, School of Health Management and Information Sciences, Iran University of Medical Sciences, Tehran, Iran. [5]Department of Medical Education and E-Learning in Medical Education, Smart University of Medical Sciences, Tehran, Iran. [6]Department of E-Learning, Virtual Faculty, and Scientific Pole of E-Learning, Shiraz University of Medical Sciences, Shiraz, Iran. ✉email: nu.kohan@gmail.com

linking the crisis to issues such as fraud, carelessness, and general unreliability[3,4]. This issue is recognized as a multifaceted and multistakeholder problem with no single cause or solution. At its core, reproducibility is defined as the ability to validate research knowledge using established science and methodology; findings that cannot be replicated will not be considered valid knowledge. Furthermore, reproducibility remains a crucial tool for evaluating scientific claims and assessing their validity[5].

Understanding the crisis requires a clear conceptual framework. Freedman et al.[6] identify three key aspects of reproducibility: reproducibility of methods (requiring detailed explanations for experts to accurately reproduce results), reproducibility of results (the technical replication of an experiment), and inferential reproducibility (determining if a reanalysis or replication yields qualitatively similar conclusions)[6].

Despite being an integral part of scientific research, many fields have experienced a decline in reproducibility in recent years[7]. Although scientific research is increasingly reliant on reproducible outcomes, many pivotal studies across multiple disciplines have not been properly replicated. An analysis of the top 100 journals in education research found that 0.13% of publications described reproducibility projects[8]. Studying 250 psychology articles published between 2014 and 2017 revealed that 5% discussed reproducibility efforts[9]; however, social sciences only mentioned reproducibility attempts 1% of the time[10]. This crisis threatens research, wasting resources, retarding hindering knowledge progression, and undermining scientific journals' credibility[11].

In medical education, the reproducibility of research results is crucial, as validated findings serve as guiding principles for curriculum development, teaching methods, and student evaluation methods[12] Evidence-based decisions in this field are essential for optimizing resource utilization and ensuring improved outcomes for faculty, learners, and health systems. Education strategies based on flawed evidence can compromise student learning and lead to the misallocation of resources toward ineffective initiatives. Moreover, a finding that has not been duly validated damages research credibility and can negatively affect health care systems overall[13,14]. While significant attention has been given to the reproducibility issue in biomedical research, comparatively less emphasis has been placed on medical education research[15-18].

This disparity highlights a critical research gap. Medical education research often lacks standardized methodologies, making replication difficult. Unlike controlled biomedical studies, this field frequently employs qualitative data and context-specific interventions[19].

Reproducibility is further complicated by the unique complexities inherent in the setting, including diverse student demographics, ethical constraints, environmental factors, and methodological issues such as inconsistent measurement tools. Consequently, results, which are often self-reported, are frequently prone to bias[20]. Recognizing the need for evidence-based improvement, this qualitative study's primary objective is to explore and interpret the various factors affecting reproducibility, identify the underlying barriers, and propose specific solutions for enhancing quality and reproducibility in Iranian medical educational research.

## Materials and methods
### Study design
This study utilized a qualitative design within an interpretive paradigm to deeply investigate the reproducibility crisis in Iranian medical education. To analyze the interview data, we employed Conventional Content Analysis based on the framework established by Graneheim and Lundman[21]. The selection of this specific inductive approach was imperative for the study's objectives.

Graneheim and Lundman's method offers a systematic rigor specifically designed for health and education contexts, capable of revealing complex phenomena through both manifest (explicit) and latent (underlying) content. This dual capability was essential for capturing the nuanced, context-dependent meanings in our experts' perspectives, ensuring that the findings remained grounded in the data rather than being limited by pre-existing theoretical biases.

### Participants
A total of twenty-four medical education experts affiliated with Iranian universities of medical sciences were recruited using purposive sampling. Participants were selected based on their specific familiarity with the concept of reproducibility. Inclusion criteria mandated a minimum of five years of experience in medical education research and scholarship, along with a record of at least ten published papers or projects in the field. Participants who did not complete the full interview process or were unavailable for follow-up were excluded from the study.

The participants in this study were aged between 35 and 44 years, with the sample exhibiting a gender distribution of 62.5% women and 37.5% men. Their demographic characteristics are summarized in Table 1.

### Data collection
Data were collected through in-depth, semi-structured interviews to gain a comprehensive understanding of participants' experiences. Before each interview, the study objectives were clearly explained, confidentiality was emphasized, and informed consent (written or verbal) was obtained. Depending on participants' preferences and convenience, interviews were conducted either face-to-face or via online platforms in a quiet setting to ensure privacy.

The interview guide was developed to encourage participants to express their views freely on research reproducibility. Each session began with general introductory questions to establish rapport, followed by the main questions. The central question was: "In your opinion, what factors influence the reproducibility of research in medical education?" In line with qualitative research traditions, the sequence of questions remained flexible to follow the flow of the participants' narratives. Probing questions were used extensively to elicit deeper insights and uncover underlying meanings in participants' perceptions, such as: "What challenges do you face in ensuring research reproducibility?", "What strategies could enhance reproducibility?", and "What actions should medical educators take to ensure reproducible research reporting?".

| Feature | Count | Percentage |
|---|---|---|
| Gender | | |
| Males | 9 | 37.5% |
| Females | 15 | 62.5% |
| Age | | |
| Under 35 years | 3 | 12.5% |
| Between 35 and 44 | 10 | 41.7% |
| Between 45 and 54 | 9 | 37.5% |
| Over 54 years | 2 | 8.3% |
| Years of Experience | | |
| Under 10 years | 3 | 12.5% |
| Between 10 and 15 | 7 | 29.2% |
| Between 16 and 20 | 8 | 33.3% |
| Over 20 years | 6 | 25% |
| University | | |
| Iran University of Medical Sciences | 5 | 20.8% |
| Tehran University of Medical Sciences | 6 | 25% |
| Shahid Beheshti University of Medical Sciences | 4 | 16.7% |
| Mashhad University of Medical Sciences | 3 | 12.5% |
| Shiraz University of Medical Sciences | 4 | 16.7% |
| Tabriz University of Medical Sciences | 2 | 8.3% |

**Table 1**. The demographic characteristics of the participants.

Data collection involved in-depth, semi-structured interviews, lasting approximately 60–90 min per participant. All interviews were recorded and subsequently transcribed verbatim to ensure the preservation of the original data complexity for subsequent analysis.

We employed a simultaneous and iterative process of data collection and analysis, which continued until data saturation was achieved, ensuring the generation of a comprehensive and rich dataset. Data were formally considered saturated when several consecutive interviews yielded no new codes or categories relevant to the research questions, and when subsequent interviews primarily served to elaborate or confirm existing themes rather than generate novel conceptual insights.

### Data analysis

Data Analysis Data analysis was conducted using the qualitative conventional content analysis approach developed by Graneheim and Lundman[21]. This inductive method was selected to systematically interpret the data through a process of decontextualization and recontextualization. Initially, the corresponding author transcribed the interviews verbatim. The research team then read the transcripts multiple times to obtain a sense of the whole and achieve immersion in the data. Following this preparatory phase, a line-by-line analysis was performed to identify "meaning units"—words, sentences, or paragraphs containing information relevant to the study's aim.

In the subsequent procedural steps, these meaning units were "condensed" to shorten the text while preserving the core content, and then abstracted and labeled with codes. Through a process of constant comparison, codes regarding similar subjects were grouped into sub-categories and categories, which represented the "manifest" content (the visible components) of the data. Finally, through deep interpretation and reflection on the underlying meanings across categories, themes were formulated to capture the "latent" content of the phenomenon (Fig. 1).

### Rigor

Trustworthiness to ensure the rigor and trustworthiness of the study, the four criteria proposed by Lincoln and Guba[22]—credibility, dependability, confirmability, and transferability—were strictly applied. Credibility was established through prolonged engagement with the data and member checking, where a summary of the coded findings was shared with participants to confirm that the results accurately reflected their experiences. Dependability was ensured through peer debriefing and external audit, where the coding process and initial findings were reviewed by two external qualitative researchers and two doctoral candidates. Confirmability was achieved by maintaining a detailed audit trail, documenting all analytical steps and decisions to allow external tracing of the research path. Finally, Transferability was facilitated by employing maximum variation sampling across participant demographics and providing a thick description of the Iranian medical education context, allowing readers to judge the applicability of the findings to similar settings (Fig. 2).

### Ethical considerations

This study received ethical approval from the Ethics Committee of Tehran University of Medical Sciences (Ethics Code: IR.TUMS.Medicine.REC.1402.515) prior to the commencement of data collection. All participants were fully briefed on the aims, procedures, and potential implications of the research to ensure their complete
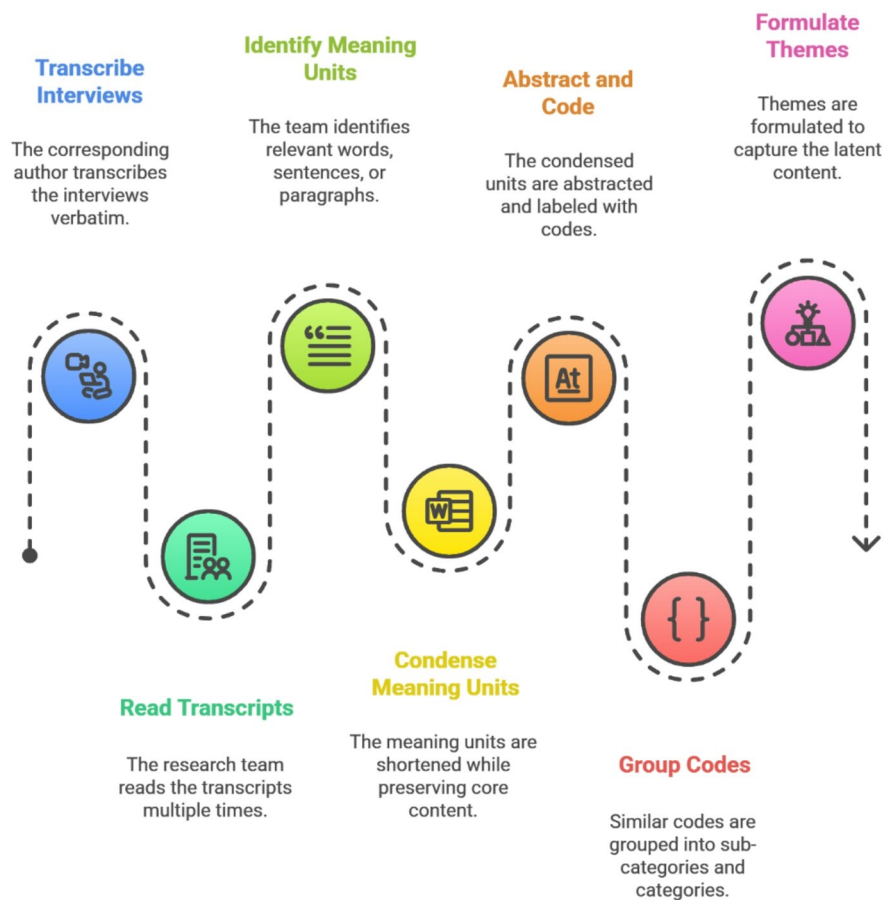
**Transcribe Interviews**

The corresponding author transcribes the interviews verbatim.

**Identify Meaning Units**

The team identifies relevant words, sentences, or paragraphs.

**Abstract and Code**

The condensed units are abstracted and labeled with codes.

**Formulate Themes**

Themes are formulated to capture the latent content.

**Read Transcripts**

The research team reads the transcripts multiple times.

**Condense Meaning Units**

The meaning units are shortened while preserving core content.

**Group Codes**

Similar codes are grouped into sub-categories and categories.

**Fig. 1**. From meaning unit to theme: steps in the qualitative content analysis.

**Transferability**

4 — Facilitating applicability with maximum variation sampling and thick description.

**Confirmability**

3 — Maintaining transparency with a detailed audit trail.

**Dependability**

2 — Ensuring consistency through peer debriefing and external audit.

**Credibility**

1 — Establishing trust through prolonged engagement and member checking.
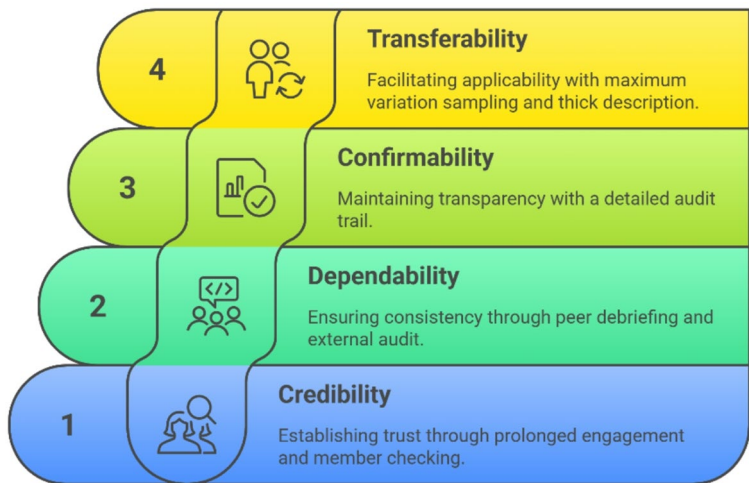
**Fig. 2**. Application of Lincoln and Guba's[22] Criteria for Trustworthiness.

understanding. Written or verbal informed consent was obtained from all participants before their inclusion in the study.

Participation was entirely voluntary, and all respondents were unequivocally assured of their right to withdraw at any point without penalty or consequence. Confidentiality and anonymity were strictly maintained throughout the research process. All interview transcripts and data were anonymized immediately upon transcription, and access to the raw data was limited exclusively to the research team. This study adhered to the ethical principles governing research involving human subjects, including the relevant guidelines outlined in the

| Theme (Latent content) | Categories (Manifest content) | Sub-categories (Manifest content) |
|---|---|---|
| 1. Factors Affecting the Reproducibility Crisis | 1.1 Research Methodology | Choosing appropriate sample size, Accurate study design, Use of reliable measurement tools |
| | 1.2 Biases | Publication bias, Sample selection bias, Reporting bias |
| | 1.3 Contextual Factors | Pressure to publish, Research funding |
| 2. Consequences of the Reproducibility Crisis | 2.1 Influence on Educational Decisions | Curriculum Design, Choosing Teaching Methods, Learning Assessment |
| | 2.2 Influence on the Progress of Science | Slowing Down the Progress of Science, Decreased Trust in Scientific Evidence |
| 3. Solutions to Deal with the Reproducibility Crisis | 3.1 Improving Research Methodology | Teaching Research Methodology, Encouraging the Use of Appropriate Statistical Methods, Emphasis on Transparency in Reporting |
| | 3.2 Changing Research Culture | Encourage Publication of Negative Results, Reduced Pressure for Publication, Increasing International Cooperation |
| | 3.3 Strengthening Research Supervision | Improving the Peer Review Process, Creating Databases for Pre-registration of Research |

**Table 2**. Overview of the analytical process: categories, sub-categories, and emergent themes.

| Sub-category | Representative quotations (Participant) |
|---|---|
| Choosing Appropriate Sample Size | *"Choosing the right sample size not only helps the validity of the results but also makes it possible to generalize the findings to larger populations." (P7)* |
| | *"In many studies, ignoring accurate sample size calculations can lead to serious biases in data interpretation." (P3)* |
| Accurate Study Design | *"The design of the study should be such that all the intervening variables are carefully controlled so as not to reduce the validity of the results." (P12)* |
| | *"A strong research design guarantees results that are reproducible and can be relied upon." (P1)* |
| Use of Reliable Measurement Tools | *"Genuine validity and reliability of the measuring instruments are part of any scientific investigation; thus, they need reflection in due caution." (P11)* |
| | *"Wrong tools may produce skewed results and misinterpretation of information." (P6)* |

**Table 3**. Participant quotations on factors: research methodology.

| Sub-category | Representative quotations (Participant) |
|---|---|
| Publication Bias | *"The tendency of journals to publish optimistic results can create a false impression of the rigor of research and lead to the erosion of the credibility of science." (P4)* |
| | *"Not publishing negative results is not only responsible for distorting the scientific literature but also diminishes public confidence in research." (P20)* |
| Sample Selection Bias | *"Selecting a sample that is not truly representative of the target population can lead to unreliable results and misinterpretations." (P10)* |
| | *"Illusion of representation in sampling is one of the major challenges that must be addressed in educational research." (P17)* |
| Reporting Bias | *"Selective reporting of results can mislead researchers and policymakers and jeopardize scientific credibility." (P5)* |
| | *"Transparency in reporting results is a key factor in maintaining credibility and trust in research results." (P25)* |

**Table 4**. Participant quotations on factors: biases.

Declaration of Helsinki. This specific research was conducted as a component of a project investigating research misconduct in medical education.

## Results

The qualitative conventional content analysis of participant interviews resulted in the identification of three major themes that capture the latent content of the data: factors affecting the reproducibility crisis, consequences of the reproducibility crisis, and solutions to deal with the reproducibility crisis. These themes, along with their respective categories and sub-categories (manifest content), are summarized in Table 2.

## Theme: factors affecting the reproducibility crisis

This theme encompasses the systemic issues within research practice and the academic environment identified by participants as contributors to the reproducibility crisis. The contributing factors were grouped into three main categories: research methodology, biases, and contextual factors.

### Research methodology

Participants highlighted specific methodological weaknesses that compromise the reliability of research findings, including failures in choosing appropriate sample size, ensuring accurate study design, and the use of reliable measurement tools. Statistical errors are also implicit methodological failures addressed by the solutions later. Representative participant insights reflecting these concerns are detailed in Table 3.

### Biases

Participants pointed to several forms of bias that distort scientific literature, specifically: Publication Bias (favoring positive results), Sample Selection Bias (non-representative sampling), and Reporting Bias (selective reporting or exaggeration of findings). Relevant participant quotations are presented in Table 4.

| Sub-category | Representative quotations (Participant) |
|---|---|
| Pressure to Publish Articles | *"The pressure from departments and promotion to produce more articles can affect the quality of research and lead to serious crises in science." (P23)* |
| | *"Research quality in departments should be evaluated based on scientific principles and not the pressures caused by publications." (P16)* |
| Research Funding | *"Financial resources can significantly influence the direction of research and lead to conflicts of interest." (P21)* |
| | *"Transparency in research funding is the key to reducing biases and maintaining scientific credibility." (P18)* |

**Table 5.** Participant quotations on factors: contextual factors.

| Sub-category | Representative quotations (Participant) |
|---|---|
| Curriculum Design | *"Using weak evidence in curriculum design can lead to adverse outcomes and inefficiencies in education." (P24)* |
| | *"Curriculums should be designed based on valid and robust evidence to help improve the quality of education." (P15)* |
| Choosing Teaching Methods | *"The adoption of weak evidence-based teaching methods can lead to inadequate student learning and a reduction in the quality of education." (P6)* |
| | *"Teaching methods should be selected based on valid research to help improve learning outcomes." (P17)* |
| Learning Assessment | *"Using flawed assessment tools may lead to false conclusions about learning and diminish the validity of the system of education." (P9)* |
| | *"Evaluation must be based on valid and reliable criteria to provide accurate results." (P12)* |

**Table 6.** Participant quotations on consequences: educational decisions.

| Sub-category | Representative quotations (Participant) |
|---|---|
| Slowing Down the Progress of Science | *"Wasting resources and time can seriously slow down the progress of science and lead to doubt about research findings." (P25)* |
| | *"The progress of science is strongly influenced by the quality of research and the validity of the results." (P19)* |
| Decreased Trust in Scientific Evidence | *"A reduction in public trust in medical research can have serious consequences for public health and undermine the credibility of science." (P14)* |
| | *"Transparency and quality of research is the key to increasing trust in scientific evidence." (P20)* |

**Table 7.** Participant quotations on consequences: progress of science.

### Contextual factors

The pressures of the academic system were identified as external drivers, including intense pressure to publish articles (quantity over quality) and the potential for conflicts of interest from research funding. Participant comments highlighting these contextual pressures are shown in Table 5.

## Theme: consequences of the reproducibility crisis

This theme details the negative impacts of non-reproducible research on educational practice and the public's perception of science.

### Influence on educational decisions

Unreliable evidence compromises the decision-making process concerning curriculum design, choosing teaching methods, and learning assessment, ultimately reducing the quality of education. Table 6 provides participant perspectives on these consequences for educational decisions.

### Influence on the progress of science

The crisis creates systemic problems that impede scientific advancement by slowing down the progress of science (wasted time and resources) and causing decreased trust in scientific evidence among the public. The related participant quotations are found in Table 7.

## Theme: solutions to deal with the reproducibility crisis

Participants proposed concrete actions grouped into three major categories that directly address the factors and consequences identified: improving research methodology, changing research culture, and strengthening research supervision.

### Improving research methodology

These solutions directly address methodological weaknesses through teaching research methodology, encouraging the use of appropriate statistical methods (to counter analysis errors), and strong emphasis on transparency in reporting. Participant suggestions for improving methodology are presented in Table 8.

### Changing research culture

Addressing the contextual and bias-related factors requires fundamental changes such as promoting the encouragement of publication of negative results (to counter publication bias), reduced pressure for publication,

| Sub-category | Representative quotations (participant) |
|---|---|
| Teaching Research Methodology | *"Providing comprehensive training in research design and implementation can help improve research quality and increase reproducibility." (P11)* |
| | *"Research should be conducted based on scientific principles and with a precise methodology in order to have high credibility." (P10)* |
| Encouraging the Use of Appropriate Statistical Methods | *"Using advanced statistical software can help improve data analysis and reduce biases." (P22)* |
| | *"Teaching statistical methods and data analysis can help reduce research errors." (P1)* |
| Emphasis on Transparency in Reporting | *"Transparency in reporting results is a key factor in maintaining credibility and trust in research results." (P3)* |
| | *"Investigations should be reported in such a way that they can be reconstructed and re-examined." (P21)* |

**Table 8.** Participant quotations on solutions: improving research methodology.

| Sub-category | Representative quotations (Participant) |
|---|---|
| Encourage Publication of Negative Results | *"Publication of negative results should be considered as an ethical necessity in educational research." (P16)* |
| | *"Negative results should be given as much attention as positive results to avoid research biases." (P8)* |
| Reduced Pressure for Publication | *"Changing the system of evaluating the performance of researchers can help reduce the pressure to publish and improve the quality of research." (P25)* |
| | *"Research should be evaluated based on quality, not quantity, so as not to reduce scientific credibility." (P13)* |
| Increasing International Cooperation | *"International cooperation can help improve the quality of research and knowledge exchange between researchers." (P19)* |
| | *"Knowledge exchange between researchers from different countries can help advance science and reduce biases." (P4)* |

**Table 9.** Participant quotations on solutions: changing research culture.

| Sub-category | Representative quotations (Participant) |
|---|---|
| Improving the Peer Review Process | *"Expert reviewers inform the process of peer review and may contribute to improving the quality of research and reducing bias." (P5)* |
| | *"The peer review process must be designed to maintain scientific credibility and avoid bias in the results." (P6)* |
| Creating Databases for Pre-registration of Research | *"Pre-registration databases can be established that allow studies to be pre-registered, thereby promoting transparency and reducing bias in science." (P10)* |
| | *"Pre-registration is a step in the right direction in terms of scientific transparency and, especially, validity of results." (P18)* |

**Table 10.** Participant quotations on solutions: strengthening research supervision.

and increasing international cooperation. These proposed cultural changes are supported by the quotations in Table 9.

### Strengthening research supervision

Solutions focusing on oversight include improving the peer review process and creating databases for pre-registration of research (to enhance transparency and counter reporting bias). Table 10 contains the relevant participant quotations.

The primary findings, synthesizing the key factors, consequences, and solutions identified in the qualitative analysis, are summarized visually in Fig. 3.

## Discussion

There are several dimensions to the reproducibility crisis in medical education research, which are highlighted in this study. Research reproducibility and validity are explained by three main themes and eight categories.

### First theme: factors affecting reproducibility crisis

This study identified three categories of factors: research methodology, biases, and contextual factors.

The first category, research methodology, is a crucial factor in ensuring reproducibility, confirming the findings of Hildebrandt and Prenoveau[23] and Klein[24]. Our findings highlight that achieving reproducibility depends on three specific categories: choosing appropriate sample size, accurate study design (a well-designed study), and the use of reliable measurement tools (well-defined measurements). Furthermore, participants implicitly indicated that flaws in statistical and data analysis methods represent methodological failures. This issue is exacerbated by the over-reliance on single metrics like the p-value and the resultant misuse of statistical methods, which the wider literature identifies as a primary driver of non-reproducibility[25,26].

Likewise, Wichman et al.[27] observe that Clinical and Translational Research (CTR) must uphold rigor and reproducibility, arguing that studies ought to be adequately collected, designed, and analyzed (addressing the statistical element). To overcome the challenges associated with each phase of CTR, tailored approaches are required. A rigorous scientific approach, transparency, and interdisciplinary collaboration can contribute significantly to the advancement of clinical research[27].

**Fig. 3**. Reproducibility crisis in medical education research.

The second category, biases, seriously compromises the quality of research findings. Our study strongly emphasizes three forms of bias. Firstly, publication bias (the tendency to prefer statistically significant findings) skews actual effect sizes and overestimates findings, a discrepancy also reported by Johnson et al.[28] in psychology research, where only 36% of replications were significant compared to 97% of originals. Secondly, sample selection bias (non-representative inclusion), which is compounded by the personalization of online search engines, threatens the completeness and validity of systematic reviews, as noted by Ćurković and Košec[29]. Thirdly, reporting bias (selective reporting or exaggeration) is a critical issue that compromises scientific integrity. The credibility of meta-analyses is similarly challenged by publication bias, subjective inclusion criteria, and variables in interpretation, as demonstrated by Lakens, Hilgard, and Staaks[30]. Blanco-Pérez et al.[31] and Mohyuddin et al.[32] agree that publication and selection bias are significant threats. Therefore, to overcome these challenges, researchers must transparently report results and adhere to standardized reporting guidelines, allowing data

sharing and preregistering their analysis plans, which, as Simonsohn[33] suggests, increases confidence and reduces bias.

The identified biases, including publication bias, sample selection bias, and reporting bias, must be analyzed within the specific institutional and systemic pressures of the Iranian academic environment. The pervasive "publish or perish" culture, driven by strict institutional promotion and ranking requirements, significantly contributes to the perpetuation of these biases. This pressure often forces researchers toward selective reporting (reporting bias) and the pursuit of statistically significant, novel, or positive findings that are more likely to be accepted by local or international journals (publication bias). Furthermore, constraints on funding and access to diverse populations can lead to non-probabilistic or convenience sampling methods, exacerbating sample selection bias. Addressing the reproducibility crisis thus requires not only methodological training but also a fundamental re-evaluation of institutional policies that prioritize quantity over research quality and rigor[34].

The third category, contextual factors, points to systemic pressures. Our findings highlight two key drivers. Scientists may be pressured to publish research quickly which can compromise the quality of their research. Kearney et al.[35] confirm that researchers under time pressure may neglect data accuracy. Additionally, our study explicitly revealed that the influence of research funding on research direction and the potential for conflicts of interest are significant contextual factors that introduce bias. To maintain high standards, it is essential to foster a research environment that prioritizes quality over quantity.

### Second theme: consequences of the reproducibility crisis

The consequences of the reproducibility crisis are two-fold, affecting both educational decisions and scientific progress.

The first consequence is the influence on educational decisions. Non-reproducible research may result in poorly conceived curricula (flawed curriculum design) and worse teaching methods (choosing teaching methods). Baker[36] affirms that only evidence that can be replicated and verified is essential for making an informed educational decision. Abid et al.[37] and Ellaway[38] concur that transparency and rigorous research are essential for high-quality health profession education. Our study further identified that reliance on unreliable research leads to the use of flawed tools for learning assessment, thereby diminishing the validity of the entire educational system. This issue is compounded by the fact that even basic steps like literature search strategies are often non-reproducible, lacking essential details like Boolean operators or search dates, as Maggio et al.[39] demonstrated.

The second consequence, influence on the progress of science, primarily involves two impacts. Firstly, non-reproducible research directly results in the slowing down the progress of science due to wasted resources and time. Secondly, and most critically, lack of reproducibility leads to decreased trust in Scientific Evidence among the public. Freese et al.[40], Wingen et al.[25], Mede et al.[41], and Nosek et al.[26] all confirm that this loss of confidence hinders scientific advancement and destroys scientific credibility. Auspurg et al.[42] also state that lack of reproducibility destroys confidence in scientific investigations. Therefore, researchers must improve the quality and transparency of their research to rebuild public trust.

### Third theme: solutions to deal with the reproducibility crisis

Our results indicate that the most important strategies fall under improving research methodology, changing research culture, and strengthening research supervision.

The first solution category, improving research methodology, is supported by previous studies and includes developing comprehensive programs for teaching research methodology and promoting the use of appropriate statistical methods to address data analysis flaws. This solution directly addresses the implicit methodological failure noted by participants, specifically the over-reliance on simplistic metrics, which requires a shift toward rigorous statistical approaches that enhance robustness and transparency[25,26]. Furthermore, emphasis on transparency in reporting of results, including detailed descriptions of methods and data, is necessary. As mentioned by Nosek et al.[26], transparency adds to the credibility of findings, fostering high public trust. Providing more detailed descriptions of research methods, releasing code and data, and preregistration of studies will enhance replicability and result in more informed clinical practice.

The second solution category, changing research culture, is essential to address systemic pressures. Key strategies include encouraging the publication of negative results (to counter publication bias) and implementing policies for reduced pressure for publication. Hail et al.[43], Baker[36], and Kedron et al.[44] emphasize the importance of prioritizing quality over quantity in the research environment. Crucially, the lack of a supportive environment necessitates institutionalizing robust Mentorship Programs to guide junior faculty through rigorous research protocols[45–47]. Our study further suggests that increasing international cooperation is a viable strategy for quality enhancement and knowledge exchange.

The final proposed solution category—Strengthening Research Supervision—focuses on enhancing oversight mechanisms to ensure research reliability. Participants specifically recommended improving the peer review process for rigor and, more fundamentally, the mandatory creation of databases for pre-registration[33]. This practice is strongly promoted as a critical means to increase transparency, reduce selective reporting bias (as highlighted by Simonsohn), and curb questionable research practices. Furthermore, a core component of enhanced supervision is demanding transparent and accurate reporting of methods and results. This entails the mandatory adoption of guidelines promoted by the EQUATOR Network (e.g., CONSORT, STROBE), as enhancing reporting standards is a fundamental and necessary step toward achieving replicability[48,49].

To make these proposed solutions immediately actionable, it is essential to draw upon successful models implemented globally. For instance, the recommendation to improve statistical methodologies can be concretely realized by adopting practices promoted by the Center for Open Science (COS) and initiatives like the TOP (Transparency and Openness Promotion) Guidelines, which compel researchers to preregister study protocols

| Identified problem (challenge/bias) | Linked theme/category | Proposed solution (actionable recommendation) | International best practice example (Ref.) |
|---|---|---|---|
| Weak Statistical Analysis | Research Methodology | Implement mandatory training in advanced statistical reporting and open science tools | TOP Guidelines (Preregistration)[50] |
| Publication Bias | Biases | Reform institutional incentive structures to reward replication and null-findings | Use of Registered Reports mechanism[53] |
| Lack of Transparency in Data | Contextual Factors | Mandate data sharing or archiving policies according to FAIR principles | Open Data Repositories (e.g., Figshare, Zenodo)[54] |
| Limited Collaboration | Strengthening Supervision | Establish multi-site collaborative research networks and mentorship programs | Successful consortia models (e.g., NIH Common Fund)[55] |

**Table 11**. Summary of identified reproducibility challenges, corresponding actionable solutions, and international best practice examples.

and analytic plans[50]. The successful adoption of preregistration in fields like psychology and economics has demonstrably reduced questionable research practices and enhanced the credibility of findings. Furthermore, promoting research transparency can be significantly strengthened by integrating FAIR (Findable, Accessible, Interoperable, Reusable)[51] data principles in institutional research policies. Exemplar models, such as the widespread adoption of standardized reporting guidelines like CONSORT (for trials) and STARD (for diagnostic accuracy)[52], have markedly improved the completeness and clarity of published reports across many international journals. By integrating these tangible, proven examples into training and institutional oversight, Iranian medical universities can effectively bridge the gap between aspirational solutions and practical, impactful reforms.

A summary of the identified challenges and their corresponding actionable solutions, along with international examples, is provided in Table 11.

The reproducibility challenges identified in this study are inevitably shaped by the institutional characteristics of Iranian medical education. Research integrity and practice are influenced by a system marked by centralized governance, hierarchical academic structures, and performance metrics that strongly prioritize publication volume. These systemic factors dictate how research quality and reproducibility are understood and enacted. Critically, however, the core concerns identified here—such as methodological flaws, systemic biases, and the prevailing pressure on researchers—align closely with the global discourse on the reproducibility crisis[43].

A comparative analysis with international studies from regions like North America and Europe reveals a shared necessity for targeted training, robust institutional policies, and community-led initiatives to reinforce reproducible research practices in the health sciences[4]. This convergence suggests that while the determinants of irreproducibility are universally acknowledged, their specific manifestation is filtered through local cultural, ethical, and organizational contexts. Therefore, effective intervention requires context-sensitive strategies. These strategies must adapt international best practices (e.g., transparent data sharing and preregistration) to the realities of Iranian medical universities, particularly by reforming incentive structures to reward quality, transparency, and replication rather than prioritizing volume alone.

## Conclusion

This qualitative study provides a comprehensive, multi-dimensional understanding of the reproducibility crisis within medical education research. The findings, based on conventional content analysis, clearly delineate the systemic threats and necessary remedies to address this phenomenon. The reproducibility crisis is driven by three main factors: methodological flaws (e.g., inadequate sample size/design and unreliable measurement tools); various Biases (specifically publication bias, selection bias, and reporting bias); and powerful contextual factors (e.g., pressure to publish and conflicts of interest from funding). The consequences of these factors are serious, leading to a reduction in the quality of educational decisions (flawed curriculum and assessment) and a fundamental loss of Public Trust in scientific evidence. Crucially, the study identifies robust solutions across three strategic domains: improving research methodology (e.g., training and appropriate statistical methods), changing academic culture (e.g., encouraging negative results and reduced publication pressure), and strengthening research supervision (e.g., pre-registration and improved peer review). Successfully implementing these comprehensive strategies is paramount to restoring scientific rigor and credibility in the field of medical education.

## Limitations

This study presents several limitations inherent in its design and execution that should be carefully considered when interpreting the findings. As a qualitative inquiry conducted through interviews within a highly specific geographic and institutional context, the results are intrinsically context-dependent and are not intended for statistical generalization across the global medical education community. Nevertheless, we employed a maximum variation sampling strategy and provided a rich, contextualized description of the setting to enhance the transferability of the findings, allowing readers to evaluate their relevance to similar contexts in medical education.

Additionally, due to the sensitive nature of the topic (research integrity and non-reproducible practices), there is an unavoidable potential for social desirability bias. Participants may have consciously or subconsciously aligned their responses with professional and academic norms, potentially understating the true extent of challenges related to research rigor. Furthermore, this study was primarily designed to identify the nature and

existence of the relevant factors and consequences; it did not quantitatively measure their prevalence in the population or the strength of the relationships between the identified variables.

### Research suggestions

Future research must primarily focus on quantitative expansion and empirical evaluation. Initially, studies should transition to large-scale quantitative methods (e.g., international surveys) to accurately measure the global prevalence of the identified factors and consequences. Concurrently, comparative studies are needed to distinguish between universal challenges and those that are context-specific to different academic systems.

Crucially, intervention studies with longitudinal designs are required to empirically evaluate the effectiveness of proposed solutions, such as mandatory advanced training and the institutional adoption of Pre-registration Databases. Finally, given the digital landscape, research must explore the impact of digital tools and search biases on reporting bias and selection bias in evidence synthesis.

## Data availability

The datasets generated and/or analyzed during the current study are available from the corresponding author on reasonable request but are not publicly available due to privacy/ethical restrictions.

## References

1. Yom, S. S. et al. Evaluating the generalizability and reproducibility of scientific research. *Int. J. Radiat. Oncol. Biol. Phys.* **113**, 1–4 (2022).
2. Cobey, K. D. et al. Biomedical researchers' perspectives on the reproducibility of research. *PLoS Biol.* **22**, e3002870 (2024).
3. Uchida, S. et al. Research reproducibility and preventing fraud. *Front. Media SA* **9**, 97946 (2022).
4. Resnik, D. B. & Shamoo, A. E. Reproducibility and research integrity. *Account. Res.* **24**, 116–123 (2017).
5. Abid, M. N., Malik, A. & Sarwar, S. Research reproducibility ethics of scientific research in higher education. *J. Educ. Res. Soc. Sci. Rev.* **3**, 37–50 (2023).
6. Freedman LP, Venugopalan G, Wisman R. Reproducibility2020: Progress and priorities. F1000Research 6 (2017).
7. Fanelli, D. Is science really facing a reproducibility crisis, and do we need it to?. *Proc. Natl. Acad. Sci. USA* **115**, 2628–2631 (2018).
8. Makel, M. C. & Plucker, J. A. Facts are more important than novelty: Replication in the education sciences. *Educ. Res.* **43**, 304–316 (2014).
9. Hardwicke, T. E. et al. Estimating the prevalence of transparency and reproducibility-related research practices in psychology (2014–2017). *Perspect. Psychol. Sci.* **17**, 239–251 (2022).
10. Hardwicke, T. E. et al. An empirical assessment of transparency and reproducibility-related research practices in the social sciences (2014–2017). *R Soc. Open Sci.* **7**, 190806 (2020).
11. Cobey, K. D. *et al.* Biomedical researchers' perspectives on the reproducibility of research: A cross-sectional international survey. bioRxiv 2023.
12. Maggio, L. A., Tannery, N. H. & Kanter, S. L. Reproducibility of literature search reporting in medical education reviews. *Acad. Med.* **86**, 1049–1054 (2011).
13. Sampson, M., Horsley, T. & Doja, A. A bibliometric analysis of evaluative medical education studies: Characteristics and indexing accuracy. *Acad. Med.* **88**, 421–427 (2013).
14. Avila, M. J. & Rodriguez-Restrepo, A. The importance of research in undergraduate medical education. *Medwave* **14**, e6032 (2014).
15. Bustin, S. A. The reproducibility of biomedical research: Sleepers awake!. *Biomol. Detect. Quantif.* **2**, 35–42 (2014).
16. McIntosh, L. D. et al. Repeat: a framework to assess empirical reproducibility in biomedical research. *BMC Med. Res. Methodol.* **17**, 143 (2017).
17. Brito, J. J. *et al.* Recommendations to enhance rigor and reproducibility in biomedical research. Gigascience 9 (2020).
18. Iqbal, S. A. et al. Reproducible research practices and transparency across the biomedical literature. *PLoS Biol.* **14**, e1002333 (2016).
19. Regehr, G. Trends in medical education research. *Acad. Med.* **79**, 939–947 (2004).
20. Hafler, J. P., Phatak, U. P. Mentoring for educational research skills and scholarship. In *Mentoring in Health Professions Education: Evidence-Informed Strategies Across the Continuum* 123–131 (Springer, 2022).
21. Graneheim, U. H. & Lundman, B. Qualitative content analysis in nursing research: concepts, procedures and measures to achieve trustworthiness. *Nurse Educ. Today* **24**, 105–112 (2004).
22. Lincoln, Y. S., Guba, E. G. *Naturalistic Inquiry*. (Sage, 1985).
23. Hildebrandt, H. & Prenoveau, J. M. Reproducibility of findings in psychological science. *Curr. Opin. Psychol.* **18**, 13–17 (2017).
24. Klein, O. A. The reproducibility crisis in science: A call for better standards. *J. Appl. Res. High. Educ.* **6**(2), 123–145 (2014).
25. Wingen, T. How to start a replication crisis. *Nat. Rev. Psychol.* **1**(5), 317–317 (2022).
26. Nosek, B. A. et al. Replicability, robustness, and reproducibility in psychological science. *Annu. Rev. Psychol.* **73**, 719–748 (2022).
27. Wichman, T. et al. Enhancing Rigor and reproducibility in clinical and translational research: A framework. *J. Clin. Transl. Res.* **6**(4), 500–512 (2020).
28. Johnson, V. E. et al. Prioritizing replication research in psychology. *Perspect. Psychol. Sci.* **12**(1), 114–123 (2017).
29. Ćurković, K. & Košec, J. Sample selection bias in systematic reviews and meta-analyses. *J. Evid. Based Med.* **14**(4), 303–315 (2021).
30. Lakens, D., Hilgard, J. & Staaks, J. Meta-analysis and the challenge of establishing generalizable knowledge. *Behav. Res. Methods.* **48**(3), 953–965 (2016).
31. Blanco-Pérez, E. et al. Publication and selection bias in medical research. *BMC Med. Res. Methodol.* **19**(1), 150 (2019).
32. Mohyuddin, H. et al. The impact of publication and selection bias on the findings of systematic reviews. *PLoS ONE* **16**(5), e0251789 (2021).
33. Simonsohn, U. The reproducibility revolution. *Nature* **527**(7577), 299–301 (2015).
34. Samadi, H. The replication crisis and the need to change the policy of scientific publication. *Methodol. Soc. Sci. Human.* **27**(108), 33–46 (2021).
35. Kearney, P. et al. Time pressure and its effect on data accuracy in scientific research. *Res. Policy.* **51**(6), 104523 (2022).
36. Baker, M. Reproducibility crisis. *Nature* **533**(7603), 452–454 (2016).
37. Abid, M. N., Malik, A. & Sarwar, S. Research reproducibility ethics of scientific research in higher education. *J. Educ. Res. Soc. Sci. Rev.* **3**(1), 37–50 (2023).
38. Ellaway, R. H. Reproducibility and replicability in health professions education research. Adv. Health Sci. Educ. Theory Pract. 1–6 (2024).

39. Maggio, L. A., Tannery, N. H. & Kanter, S. L. Reproducibility of literature search reporting in medical education reviews. *Acad. Med.* **86**(8), 1049–1054 (2011).
40. Freese, J., Rauf, T. & Voelkel, J. G. Advances in transparency and reproducibility in the social sciences. *Soc. Sci. Res.* **107**, 102770 (2022).
41. Mede, N. G. et al. The "replication crisis" in the public eye: Germans' awareness and perceptions of the (IR) reproducibility of scientific research. *Public Underst. Sci.* **30**(1), 91–102 (2021).
42. Auspurg, K., Brüderl, J. How to increase reproducibility and credibility of sociological research. In *Handbook of Sociological Science*. (Edward Elgar Publishing, 2022).
43. Hail, L. et al. Prioritizing quality over quantity: A proposal for research evaluation. *Res. Eval.* **31**(3), 320–330 (2022).
44. Kedron, P., Holler, J. & Bardin, S. reproducible research practices and barriers to reproducible research in geography: Insights from a survey. *Ann. Am. Assoc. Geogr.* **114**(2), 369–386 (2024).
45. Choi, A. M. K., Moon, J. E., Steinecke, A. & Prescott, J. E. Developing a culture of mentorship to strengthen academic medical centers. *Acad. Med.* **94**(5), 630–633 (2019).
46. Williams, J. S. et al. Mentoring strategies to support diversity in research-focused junior faculty: A scoping review. *J. Clin. Transl. Sci.* **7**(1), e21 (2023).
47. da Silva Souza, R. C., Bersaneti, M. D., dos Santos Yamaguti, W. P. & Baia, W. R. Mentoring in research: Development of competencies for health professionals. *BMC Nurs.* **21**(1), 244 (2022).
48. Simera, I. et al. Transparent and accurate reporting increases reliability, utility, and impact of your research: reporting guidelines and the EQUATOR Network. *BMC Med.* **8**, 24 (2010).
49. Prager, E. M. et al. Improving transparency and scientific rigor in academic publishing. *Cancer Rep. (Hoboken).* **2**(5), e1150 (2019).
50. Patarčić, I. & Stojanovski, J. Adoption of transparency and openness promotion (TOP) guidelines across journals. *Publications.* **10**(4), 46 (2022).
51. Wilkinson, M. D. et al. The FAIR Guiding Principles for scientific data management and stewardship. *Sci. Data.* **3**(1), 1–9 (2016).
52. Bossuyt, P. M., Reitsma, J. B., Bruns, D. E., Gatsonis, C. A., Glasziou, P. P., Irwig, L. M., Lijmer. J.G., Moher, D., Rennie. D. Towards complete and accurate reporting of studies of diagnostic accuracy: The STARD initiative.
53. Chambers, C. D. & Tzavella, L. The past, present and future of Registered Reports. *Nat. Hum. Behav.* **6**(1), 29–42 (2022).
54. Crespo Garrido, I. D., Gutleber, J., Loureiro García, M. Springer: The Value of an Open Scientific Data and Documentation Platform in a Global Project: The Case of Zenodo.
55. Ellis-Robinson, T. et al. Collaborative action research with diverse stakeholders: Building the Disability Champions Mentoring Network. *Career Dev. Transit. Except. Individ.* **48**(1), 21–33 (2025).

## Acknowledgements

## Author contributions

SA conceived the study design and supervised all stages. NK led data analysis, manuscript drafting, and correspondence. HH contributed to data interpretation and language editing. AV and MH participated in data collection, visualization, and critical revisions. All authors approved the final manuscript.

## Declarations

### Competing interests

The authors declare no competing interests.

### Ethical approval

This study received ethical approval from the Ethics Committee of Tehran University of Medical Sciences (Ethics Code: IR.TUMS.Medicine.REC.1402.515) prior to the commencement of data collection. All participants were fully briefed on the aims, procedures, and potential implications of the research to ensure their complete understanding. Written or verbal informed consent was obtained from all participants before their inclusion in the study.

### Consent to participate

Participation was entirely voluntary, and all respondents were unequivocally assured of their right to withdraw at any point without penalty or consequence. Confidentiality and anonymity were strictly maintained throughout the research process. All interview transcripts and data were anonymized immediately upon transcription, and access to the raw data was limited exclusively to the research team. This study adhered to the ethical principles governing research involving human subjects, including the relevant guidelines outlined in the Declaration of Helsinki. This specific research was conducted as a component of a project investigating research misconduct in medical education.

### Consent for publication

All participants in this study provided written informed consent for the publication of findings and related data.

### Additional information

**Correspondence** and requests for materials should be addressed to N.K.

**Reprints and permissions information** is available at www.nature.com/reprints.

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.