

RESEARCH ARTICLE

Creation of standardized tools to evaluate reporting in health research: Population Reporting Of Gender, Race, Ethnicity & Sex (PROGRES)

Anne M. Stey^{1*}, Mira Ghneim², Onaona Gurney³, Ariel P. Santos⁴, Rishi Rattan⁵, Egide Abahuje⁶, Archit Baskaran¹, Jeffry Nahmias⁶, Joseph Richardson⁷, Tanya L. Zakrison⁸, Zinzi D. Baily⁹, Elliott R. Haut¹⁰, Mihir Chaudhary¹⁰, Bellal Joseph¹¹, Ben Zarzaur¹², Kimberly Hendershot¹³

1 Department of Surgery, Feinberg School of Medicine, Northwestern University, Chicago, IL, United States of America, **2** R Adams Cowley Shock Trauma Center, University of Maryland School of Medicine, Baltimore, MD, United States of America, **3** Department of Surgery, New York University Langone Health, Brooklyn, NY, United States of America, **4** Department of Surgery, Texas Tech University Health Sciences Center, Lubbock, TX, United States of America, **5** DeWitt Daughtry Family Department of Surgery, University of Miami, Miami, FL, United States of America, **6** Department of Surgery, University of California, Irvine, Orange, CA, United States of America, **7** Department of African American Studies, University of Maryland, College Park, MD, United States of America, **8** Section of Trauma and Acute Care Surgery, University of Chicago, Chicago, IL, United States of America, **9** Soffer Clinical Research Center, University of Miami, Miami, FL, United States of America, **10** Department of Surgery, Division of Acute Care Surgery, Johns Hopkins University, Baltimore, MD, United States of America, **11** Department of Surgery, Division of Trauma, Critical Care, Emergency Surgery and Burns, College of Medicine, University of Arizona, Tucson, AZ, United States of America, **12** Department of Surgery, University of Wisconsin, Madison, WI, United States of America, **13** Department of Surgery, Division of Acute Care Surgery, University of Alabama Birmingham, Birmingham, AL, United States of America

* as013j@gmail.com

Abstract

Despite increasing diversity in research recruitment, research finding reporting by gender, race, ethnicity, and sex has remained up to the discretion of authors. This study developed and piloted tools to standardize the inclusive reporting of gender, race, ethnicity, and sex in health research. A modified Delphi approach was used to develop standardized tools for the inclusive reporting of gender, race, ethnicity, and sex in health research. Health research, social epidemiology, sociology, and medical anthropology experts from 11 different universities participated in the Delphi process. The tools were pilot tested on 85 health research manuscripts in top health research journals to determine inter-rater reliability of the tools. The tools each spanned five dimensions for both sex and gender as well as race and ethnicity: Author inclusiveness, Participant inclusiveness, Nomenclature reporting, Descriptive reporting, and Outcomes reporting for each subpopulation. The sex and gender tool had a median score of 6 and a range of 1–15 out of 16 possible points. The percent agreement between reviewers piloting the sex and gender tool was 82%. The interrater reliability or average Cohen's Kappa was 0.54 with a standard deviation of 0.33 demonstrating moderate agreement. The race and ethnicity tool had a median score of 1 and a range of 0–15 out of 16 possible points. Race and ethnicity were both reported in only 25.8% of studies



OPEN ACCESS

Citation: Stey AM, Ghneim M, Gurney O, Santos AP, Rattan R, Abahuje E, et al. (2023) Creation of standardized tools to evaluate reporting in health research: Population Reporting Of Gender, Race, Ethnicity & Sex (PROGRES). *PLOS Glob Public Health* 3(9): e0002227. <https://doi.org/10.1371/journal.pgph.0002227>

Editor: Behdin Nowrouzi-Kia, University of Toronto Temerty Faculty of Medicine, CANADA

Received: June 30, 2022

Accepted: July 11, 2023

Published: September 7, 2023

Copyright: © 2023 Stey et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Data Availability Statement: All data can be found in the manuscript and [supporting information](#) files.

Funding: The authors received no specific funding for this work.

Competing interests: The authors have declared that no competing interests exist.

evaluated. Most studies that reported race reported only the largest subgroups; White, Black, and Latinx. The percent agreement between reviewers piloting the race and ethnicity tool was 84 and average Cohen's Kappa was 0.61 with a standard deviation of 0.38 demonstrating substantial agreement. While the overall dimension scores were low (indicating low inclusivity), the interrater reliability measures indicated moderate to substantial agreement for the respective tools. Efforts in recruitment alone will not provide more inclusive literature without improving reporting.

Introduction

Over the last decade, there has been a concerted effort to address noticeable gaps in participant representation in health research. For instance, the National Institutes of Health (NIH)—the nation's primary medical research agency—have launched initiatives, such as “All of US,” which seek to improve the diversity of research participants [1]. Additionally, all NIH funded research must attempt to recruit diverse participants to produce results generalizable to the broader population [2]. The ultimate goal of scientific health research at large is to produce greater diversity, transparency, accessibility, and generalizability of health research data [3]. However, despite increasing diversity in recruitment and research participation, the ultimate reporting of research findings by gender, race, ethnicity, and sex has remained unspecified and up to the discretion of authors [4].

The Enhancing the QUALity and Transparency Of health Research (EQUATOR) Network is a global initiative that has devised standardized reporting guidelines for most types of biomedical research study designs (i.e. randomized trials, systematic reviews, and observational studies) [5]. The EQUATOR mission is to achieve accurate, complete and transparent reporting of all health research studies to support research reproducibility and usefulness [1]. These guidelines have been adopted by scientific journals, such as the JAMA network, and serve as a reference for scientific health research reporting [6,7]. However, no EQUATOR guidelines exist for the standardized reporting of gender, race, ethnicity, and sex.

Many published calls to action shed light on the poor reporting of gender, race, ethnicity, and sex, including one by the Eastern Association for the Surgery of Trauma (EAST) with multidisciplinary colleagues representing clinician researchers, social epidemiologists, sociologists, and medical anthropologists [8]. Poor reporting prevents better understanding of health outcomes in marginalized populations, such as immigrant, Black, and Indigenous people [9]. No study to date has suggested how to operationalize standardized reporting. This study sought to develop and then pilot tools to standardize the inclusive reporting of gender, race, ethnicity, and sex in health research.

Methods

This study used a modified Delphi approach to develop a standardized tool for the inclusive reporting of gender, race, ethnicity, and sex in health research. The subsequent tools were pilot tested on health research manuscripts in four top health research journals to determine interrater reliability of the tools.

Expert convening

Health research, social epidemiology, sociology, and medical anthropology experts were identified from the EAST professional society and beyond. Initially 30 Individuals from the

Multicenter Trials and the Equity, Quality, and Inclusion in Trauma Surgery Practice Committees were invited to join. A total of 15 experts from 11 different universities representing all regions of the United States agreed to participate. The participants and non-participants were similar in age, gender, ethnicity, sex, academic and clinical background (S1 Table).

Literature review

The aim of the literature review was to ascertain the breadth of categories for Gender, Race, Ethnicity and Sex in the English health research evidence. A literature search was performed querying Medline for the search terms “sex reporting” “gender reporting” “race reporting” and “ethnicity reporting”. There were no Mesh terms utilized. A time frame from 1960 to September 2020 was searched to chronicle the evolution of demographic categorizations over time. We limited the search terms to these four terms because of the long study period we were considering and the huge volume of articles that we chose to prioritize specificity as opposed to sensitivity of our search. A librarian assisted us in developing the search terms and pulling the articles. Only English-language peer-reviewed journal articles were included. The 15 individuals were divided into three groups of five people, one that screened and extracted data for sex and gender literature, one group that screened and extracted data for race literature, and one group that screened and extracted data for ethnicity literature. All members of each group independently screened titles, abstracts, and full text manuscripts. Discussion among all five members of each of the three groups was used to reach consensus for final full text inclusion for extraction. Reference lists from included full text publications were screened to identify other relevant literature.

Extraction and statement development

Data regarding reporting equity in full text manuscripts were extracted into a Microsoft Word document. Extracted data were ranked according to level of evidence. The level of evidence in the available literature ranged from systematic reviews to expert opinion. The entire group reviewed the document and met twice as a group to achieve a draft statement consensus document summarizing the findings and recommendations based on the 19 full-texts included in the review [8].

Tool development

The draft statement consensus document contained a series of recommendations for improved reporting equity based on the literature review [8]. Tools to assess the quality of reporting of gender, race, ethnicity, and sex in published manuscripts was devised by two individuals (AS, KH) based on a three-step modified Delphi consensus process which took place between September to December 2020 [10]. A comprehensive list of five dimensions was drafted after expert participants reviewed the existing literature. This list was iteratively reviewed using systematic progression of repeated rounds of voting to determine expert group consensus. The modified Delphi consensus process consisted of two rounds of emails and one virtual meeting. The tools quantified reporting along five dimensions: authorship inclusiveness, participant inclusiveness, nomenclature, descriptive reporting, and outcome reporting. A comprehensive list of potential components organized into these five dimensions was drafted. Each dimension specified 1–2 components. This list was iteratively reviewed using systematic progression of repeated rounds of voting to determine expert group consensus. The modified Delphi consensus process consisted of one round of emails and one virtual meeting.

Round 1. The tool was circulated via email to the 15 participants with the goal to optimize the tools' syntax. Each individual was asked to review and agree or disagree with each

component in the sex and gender tool as well as the race and ethnicity tool. Responses were gathered via email. Components required 80% agreement to be accepted or omitted. Components that did not reach 80% threshold for acceptance on the first round were adapted via participant input and redistributed in the second round.

Round 2. The same consensus method was used in this round but accomplished via a virtual meeting. Participants were encouraged to discuss the entire tools and all components until agreement was reached to retain, modify, or eliminate components. Responses were collated and analyzed and the tool, adapted based on recommendations from participants.

Round 3. The tool in its entirety was then circulated again via email. The same consensus method was used as in round 1. Final responses were analyzed and described with components reaching 80% agreement being retained in the final tools.

Pilot testing

Using the tools, seven individual from the 15 experts assessed the 66 published original research reports from December 2020 in four of the highest impact health research journals: the New England Journal of Medicine, Journal of the American Medical Association, the British Medical Journal, and the Lancet (S2 Table). Additionally, the tools were piloted utilizing the manuscripts from 19 Multicenter Trials supported by the Eastern Association for the Surgery of Trauma, the Western Trauma Association, and the American Association for the Surgery of Trauma from January 2019-January 2021 to oversample reports from a field of health research where participants and patients are disproportionately from marginalized racial and ethnic groups. There was no deliberate oversampling of original research focused on specific racial and ethnic populations (e.g., the Sister Study, Black Women's Health Study). Each published report was reviewed for gender, race, ethnicity, and sex reporting in each of five dimensions. Author inclusiveness reporting was assessed based on personal knowledge of authors, or internet web search for gender, race, ethnicity, and sex identification. Participant inclusiveness reporting was assessed based on whether multiple gender, race, ethnicity, and sex were recruited or included. Nomenclature reporting was assessed based on the utilization of specific gender, race, ethnicity, and sex categories specified by the US census [11]. Descriptive reporting was assessed based on whether the results section or descriptive tables presented data by gender, race, ethnicity, and sex composition. Outcomes reporting was assessed based on whether the results section or univariate and multivariate outcomes tables reported data for each subpopulation separately.

Two reviewers scored each study independently using the tools in blinded fashion. Each study was scored based on inclusion or exclusion of each component within the sex and gender tool (S1 Data) and the race and ethnicity tool, respectively (S2 Data). Interrater reliability (measured as percent agreement and Cohen's kappa) was calculated between the two reviewers for each component of each tool. Components with less than 90% agreement were excluded from the final tool. Interrater reliability was also assessed for each of the overall tools.

Results

Two tools emerged from the review of literature and Delphi process. The tools each spanned five dimensions for both sex and gender (Table 1) as well as race and ethnicity (Table 2). Author inclusiveness reporting assessed whether studies had diverse authorship, which may influence analysis, interpretation, and results reporting. Participant inclusiveness reporting assessed whether a broad range of participants were recruited. Nomenclature reporting assessed how granular gender, race, ethnicity, and sex were captured. By granular, we mean how detailed were gender, race, ethnicity and sex reported. Descriptive reporting assessed

Table 1. Sex and gender reporting tool generated from the delphi process.

	Dimension	Description	Points	
1	Authorship inclusiveness	Author representation of different sexes and genders	1	
2	Participant inclusiveness	Participant inclusion of different sexes	1	
		Participant inclusion of different genders	1	
		Why one or both not applicable		
3	Nomenclature	Gender	Men	1
			Women	1
			Transgender	1
			Non-binary/Non-conforming	1
		Sex	Male	1
			Female	1
			Inter-sex	1
4	Descriptive reporting	By sex	1	
		By gender	1	
5	Outcome reporting	Univariate outcomes by sex	1	
		Multivariate outcomes by sex	1	
		Univariate outcomes by gender	1	
		Multivariate outcomes by gender	1	
	Total Points		16	

<https://doi.org/10.1371/journal.pgph.0002227.t001>

whether descriptive statistics included gender, race, ethnicity, and sex composition. Outcomes reporting assessed whether univariate and multivariate outcomes were reported separately for each subpopulation. The tools did not consider the assessment of interactions, stratified or sensitivity analyses.

Table 2. Race and ethnicity reporting tool generated from the delphi process.

	Dimensions	Description	Points	
1	Author inclusiveness	Author representation of different races and ethnicities	1	
2	Participant inclusiveness	Participant inclusion of different races	1	
		Participant inclusion of different ethnicities Or why one or both not applicable	1	
3	Nomenclature	Race	Native American or Alaska Native	1
			Asian	1
			Black	1
			Native Hawaiian or Pacific Islander	1
			White	1
		Ethnicity	Hispanic/Latinx *	1
			Classification beyond Hispanic**	1
4	Descriptive reporting	By race	1	
		By ethnicity	1	
5	Outcome reporting	Univariate outcomes by race	1	
		Multivariate outcomes by race	1	
		Univariate outcomes by ethnicity	1	
		Multivariate outcomes by ethnicity	1	
	Total Points		16	

* NIH Standard Definition

** Supplemental table.

<https://doi.org/10.1371/journal.pgph.0002227.t002>

Sex and gender reporting components were each dichotomous and ranged from 0 to 1 in the 85 published manuscripts in peer-reviewed journals. The sex and gender tool had a median score of 6 and a range of 1–15 out of 16 possible points. All studies captured sex or gender in at least one dimension of the tool. However, there was repeated conflation between biologic sex and self-reported gender, and sexual orientation (not explicitly assessed in this tool), and no studies reporting both sex and gender. There was extremely low reporting of non-binary people, transgender men, transgender women, and intersex people. If transgender people were included, most studies collapsed it into one single monolithic label without differentiation between patients' specific identities within the transgender community (ex. transmasculine, transfeminine, transgender woman, transgender man, etc). Despite these challenges, the overall percent agreement between reviewers piloting the sex and gender tool was 82%. The interrater reliability or average Cohen's Kappa for all reviewers reviewing all studies with the sex and gender tool was 0.54 with a standard deviation of 0.33 demonstrating moderate agreement.

The race and ethnicity tool had a median score of 1 and a range of 0–15 out of 16 possible points. Race and ethnicity were both reported in only 25.8% of the high impact studies evaluated. Most studies that reported race reported only the largest subgroups such as White, Black, and Latinx. Despite these challenges, the overall percent agreement between reviewers piloting the tool for the race and ethnicity tool was 84%. The interrater reliability or average Cohen's Kappa for the reviewers reviewing all studies with the race and ethnicity tool was 0.61 with a standard deviation of 0.38 demonstrating substantial agreement.

Discussion

In our study, we were able to develop and pilot a tool to standardize the inclusive reporting of gender, race, ethnicity, and sex in health research. Despite recent initiatives in trying to capture inclusive study populations, there appears to be a lack of granular reporting of data for specific marginalized populations who are typically burdened disproportionately by negative health outcomes [4]. Poor reporting prevents better understanding of health outcomes in marginalized populations, such as transgender, Black, and Indigenous people [9]. There have been previous calls for more standardized reporting of gender, race, ethnicity and sex [12,13]. This study sought to develop a tool for standardized reporting of gender, race, ethnicity, and sex in health research. As such, we developed and piloted a joint reporting tool which demonstrated that recent literature had better reporting of dimensions of sex and gender than for dimensions of race and ethnicity. The interrater reliability was moderate for the sex and gender reporting and substantial for the race and ethnicity reporting tool.

The results of this study compares well with prior studies that demonstrate a great degree of heterogeneity in the reporting of gender, race, ethnicity [14], and sex [15,16]. There have been calls for improved reporting across all of health research ranging from alcohol-use disorders to cancer over the last two decades [17,18]. However, to our knowledge this is the first study that has developed and piloted of a tool to standardize gender, race, ethnicity, and sex reporting in health research. If further validated, this reporting tool could be particularly powerful upon adoption by the EQUATOR network by allowing more widespread adoption for standardized reporting of gender, race, ethnicity, and sex in all disciplines of health research.

There are several limitations to this study. First, the expert panel convened included few researchers from other disciplines and could have included more people from under-represented and marginalized groups. Although we attempted to adhere to established gender, race, ethnicity, and sex categories based on the United States Census, and the National Institutes of Health [19], we seek input to revise and make more inclusive tools to better capture our diverse

Table 3. Ethnic Ancestry Categories in the United States [19].

Afghan	Bangladeshi	Croatian	Filipino	Honduran
African	Barbadian	Cuban	Finnish	Hungarian
African American	Belgian	Cypriot (Greek and Turkish)	French	Icelandic
Albanian	Bhutanese	Czech	French Canadian	Indian
American	Brazilian	Czechoslovakian	Georgian	Indonesian
American Indian	British	Danish	German	Iranian
Arab	Cambodian	Dominican	Greek	Irish
Argentinean	Canadian	Dutch	Guatemalan	Israeli
Armenian	Chilean	Ecuadorian	Guyanese	Italian
Asian	Chinese	Ethiopian	Haitian	Jamaican
Asian Indian	Colombian	English	Hispanic	Japanese
Austrian	Costa Rican		Hmong	Korean
Laotian	Peruvian	Serbian	Trinidadian and Tobagonian	
Latin American	Polish	Sikh	Turkish	
Lithuanian	Portuguese	Sri Lankan	Ukrainian	
Malaysian	Puerto Rican	Slavic	Venezuelan	
Mexican	Romanian	Slovak	Vietnamese	
Nicaraguan	Russian	Slovene	Welsh	
Nigerian	Salvadorian	Spaniard	West Indian	
Nepali	Samoa	Swedish	Yugoslavian	
Norwegian	Scandinavian	Swiss		
Pakistani	Scottish Irish	Taiwanese		
Panamanian	Scottish	Thai		

<https://doi.org/10.1371/journal.pgph.0002227.t003>

population (Table 3). Second, these tools have not been extensively validated by raters who were not part of this panel. This will be an important future direction. Third, pilot testing with manuscripts from the four highest impact health research journals puts the findings at risk for selection bias due to bias in what is accepted for publication in high impact health journals. However, we attempted to overcome this limitation by also including manuscripts from several multicenter trials published in a variety of journals. Fourth, author names, personal knowledge of authors, and web searches were used to give points for authorship inclusiveness scoring. However, this approach would not be easily replicated in the broader literature because names and web pages may not reflect authors' gender, race, ethnic, or sex identity. We recommend that upon implementation of these tools, authors should be asked to self-identify this information. Furthermore, we encourage journal to routinely collect these data upon submission and publication to assess journal-specific author inclusivity over time. Fifth, race and ethnicity are social constructs, and our conceptualization is based on a United States-centric approach to race and ethnicity. For instance, the race and ethnicity reporting tool was focused on overall populations with known disparities in care within the United States, while recognizing that the large categories hide wide heterogeneity (e.g., different ethnicities within the Asian or Black category). We recommend that the race and ethnicity tool be contextually adapted for regions when applied internationally in order to accurately capture local ethnicities facing disparities in health and a disproportionate burden of social determinants of health inequities.

There remains a great deal of opportunity to ensure a more inclusive recruitment and enrollment of diverse and marginalized populations within health research. We believe this will translate into improved methodological analyses and reporting of important research findings. This study involved the design and piloting of a tool for standardized reporting of gender, race, ethnicity, and sex in health research. While the overall dimension scores were relatively low (indicating low inclusivity), the interrater reliability measures indicated moderate to substantial agreement for the respective tools. All the efforts in recruitment and enrollment will

not alone provide richer, more inclusive scientific health research literature without improving detailed reporting of data involving under-represented and marginalized populations.

Supporting information

S1 Table. Demographic characteristics of experts who participated in the delphi process.

Legend: This table denotes the aggregate age, gender, and ethnicity of participants of the literature review and delphi process.

(DOCX)

S2 Table. Manuscripts reviewed for piloting gender race ethnicity and sex reporting tool.

Legend: This table contains the title, journal and link to articles in top health journals used to pilot the gender race ethnicity and sex reporting tools.

(XLSX)

S1 Data. Gender and sex reporting tool data collected. Articles with gender and sex reporting rating by study team with analytic macro of concordance.

(XLSM)

S2 Data. Race and ethnicity reporting tool data collected. Legend: Articles with race and ethnicity reporting rating by study team with analytic macro of concordance.

(XLSM)

Author Contributions

Conceptualization: Anne M. Stey, Mira Ghneim, Onaona Gurney, Ariel P. Santos, Rishi Rattan, Egide Abahuje, Jeffrey Nahmias, Joseph Richardson, Tanya L. Zakrison, Zinzi D. Baily, Elliott R. Haut, Bellal Joseph, Ben Zarzaur, Kimberly Hendershot.

Data curation: Anne M. Stey, Mira Ghneim, Onaona Gurney, Ariel P. Santos, Rishi Rattan, Archit Baskaran, Kimberly Hendershot.

Formal analysis: Anne M. Stey, Mira Ghneim, Onaona Gurney, Ariel P. Santos, Rishi Rattan, Archit Baskaran, Kimberly Hendershot.

Writing – original draft: Anne M. Stey, Mira Ghneim, Onaona Gurney, Ariel P. Santos, Rishi Rattan, Egide Abahuje, Archit Baskaran, Jeffrey Nahmias, Joseph Richardson, Tanya L. Zakrison, Zinzi D. Baily, Elliott R. Haut, Mihir Chaudhary, Bellal Joseph, Ben Zarzaur, Kimberly Hendershot.

Writing – review & editing: Anne M. Stey, Mira Ghneim, Onaona Gurney, Ariel P. Santos, Rishi Rattan, Egide Abahuje, Archit Baskaran, Jeffrey Nahmias, Joseph Richardson, Tanya L. Zakrison, Zinzi D. Baily, Elliott R. Haut, Mihir Chaudhary, Bellal Joseph, Ben Zarzaur, Kimberly Hendershot.

References

1. Ilori TO, Viera E, Wilson J, Moreno F, Menon U, Ehiri J, et al. Approach to High Volume Enrollment in Clinical Research: Experiences from an All of Us Research Program Site. *Clin Transl Sci.* 2020 Jul; 13(4):685–92. <https://doi.org/10.1111/cts.12759> PMID: 32004412. PMCID: PMC7359931. Epub 2020/03/17. eng.
2. Corbie-Smith GM, Durant RW, St George DM. Investigators' assessment of NIH mandated inclusion of women and minorities in research. *Contemp Clin Trials.* 2006 Dec; 27(6):571–9. <https://doi.org/10.1016/j.cct.2006.05.012> PMID: 16843069. Epub 2006/07/13. eng.

3. Hohmann AA, Parron DL. How the new NIH Guidelines on Inclusion of Women and Minorities apply: efficacy trials, effectiveness trials, and validity. *J Consult Clin Psychol*. 1996 Oct; 64(5):851–5. <https://doi.org/10.1037//0022-006x.64.5.851> PMID: 8916610. eng.
4. Ma IW, Khan NA, Kang A, Zalunardo N, Palepu A. Systematic review identified suboptimal reporting and use of race/ethnicity in general medical journals. *J Clin Epidemiol*. 2007 Jun; 60(6):572–8. <https://doi.org/10.1016/j.jclinepi.2006.11.009> PMID: 17493512. Epub 2007/03/26. eng.
5. Thornton H. Report on EQUATOR Network launch meeting 26th June 2008 "Achieving Transparency in Reporting Health Research". *Int J Surg*. 2008 Dec; 6(6):428–31. <https://doi.org/10.1016/j.ijsu.2008.08.002> PMID: 18804424. Epub 2008/08/15. eng.
6. Catalá-López F, Alonso-Arroyo A, Page MJ, Hutton B, Ridaio M, Tabarés-Seisdedos R, et al. Reporting guidelines for health research: protocol for a cross-sectional analysis of the EQUATOR Network Library. *BMJ Open*. 2019 03; 9(3):e022769. <https://doi.org/10.1136/bmjopen-2018-022769> PMID: 30837245. PMCID: PMC6429992. Epub 2019/03/04. eng.
7. Brooke BS, Ghaferi AA, Kibbe MR. Effective Use of Reporting Guidelines to Improve the Quality of Surgical Research. *JAMA Surg*. 2021 Jun; 156(6):515–6. <https://doi.org/10.1001/jamasurg.2021.0519> PMID: 33825817. eng.
8. Nahmias J, Zakrisson TL, Haut ER, Gurney O, Joseph B, Hendershot K, et al. Call to Action on the Categorization of Sex, Gender, Race, and Ethnicity in Surgical Research. *J Am Coll Surg*. 2021 May. <https://doi.org/10.1016/j.jamcollsurg.2021.04.025> PMID: 33964401. Epub 2021/05/05. eng.
9. Haas AP, Lane A, SO/GI WGfPlo. Collecting Sexual Orientation and Gender Identity Data in Suicide and Other Violent Deaths: A Step Towards Identifying and Addressing LGBT Mortality Disparities. *LGBT Health*. 2015 Mar; 2(1):84–7. <https://doi.org/10.1089/lgbt.2014.0083> PMID: 26790023. PMCID: PMC4713015. Epub 2015/01/13. eng.
10. Dalkey N HO. An experimental application of the Delphi method to the use of experts. *Management Science*. 1963; 9(3):458–67.
11. Program NloHD. Racial and Ethnic Categories and Definitions for NIH Diversity Programs and for Other Reporting Purposes <http://grants.nih.gov/grants/guide/notice-files/NOT-OD-15-089.html2015>.
12. Clayton JA, Tannenbaum C. Reporting Sex, Gender, or Both in Clinical Research? *JAMA*. 2016 Nov; 316(18):1863–4. <https://doi.org/10.1001/jama.2016.16405> PMID: 27802482. eng.
13. Heidari S, Babor TF, De Castro P, Tort S, Curno M. Sex and Gender Equity in Research: rationale for the SAGER guidelines and recommended use. *Res Integr Peer Rev*. 2016; 1:2. <https://doi.org/10.1186/s41073-016-0007-6> PMID: 29451543. PMCID: PMC5793986. Epub 2016/05/03. eng.
14. Kaplan JB, Bennett T. Use of race and ethnicity in biomedical publication. *JAMA*. 2003 May; 289(20):2709–16. <https://doi.org/10.1001/jama.289.20.2709> PMID: 12771118. eng.
15. Mansukhani NA, Yoon DY, Teter KA, Stubbs VC, Helenowski IB, Woodruff TK, et al. Determining If Sex Bias Exists in Human Surgical Clinical Research. *JAMA Surg*. 2016 11; 151(11):1022–30. <https://doi.org/10.1001/jamasurg.2016.2032> PMID: 27551816. PMCID: PMC5142632. eng.
16. Etherington N, Wu M, Boet S. Sex/gender and additional equity characteristics of providers and patients in perioperative anesthesia trials: a cross-sectional analysis of the literature. *Korean J Anesthesiol*. 2020 Mar. <https://doi.org/10.4097/kja.19484> PMID: 32164399. Epub 2020/03/13. eng.
17. Schick MR, Spillane NS, Hostetler KL. A Call to Action: A Systematic Review Examining the Failure to Include Females and Members of Minoritized Racial/Ethnic Groups in Clinical Trials of Pharmacological Treatments for Alcohol Use Disorder. *Alcohol Clin Exp Res*. 2020 Sep. <https://doi.org/10.1111/acer.14440> PMID: 32997374. Epub 2020/09/30. eng.
18. Borno HT, Small EJ, Zhang L, DeRouen MC, Griffin A, McGuire J, et al. How current reporting practices may mask differences: A call for examining cancer-specific demographic enrollment patterns in cancer treatment clinical trials. *Contemp Clin Trials Commun*. 2019 Dec; 16:100476. <https://doi.org/10.1016/j.conctc.2019.100476> PMID: 31872154. PMCID: PMC6915794. Epub 2019/10/28. eng.
19. Kelly Mathews JP, Nicholas A, Jones, Konya Sarah, Marks Rachel, Pratt Beverly M., Coombs Julia, et al. 2015 National Content Test Race and Ethnicity Analysis Report <https://www2.census.gov/programs-surveys/decennial/2020/program-management/final-analysis-reports/2015nct-race-ethnicity-analysis.pdf>. United States Department of Commerce Economics and Statistics Administration; 2017 [1].