Supplementary Material

To Report or Not to Report on Research Ethics in Political Science and International Relations: A New Dimension of Gender-based Inequality

Eleanor Knott and Denisa Kostovicova

Contents

1 Search Terms

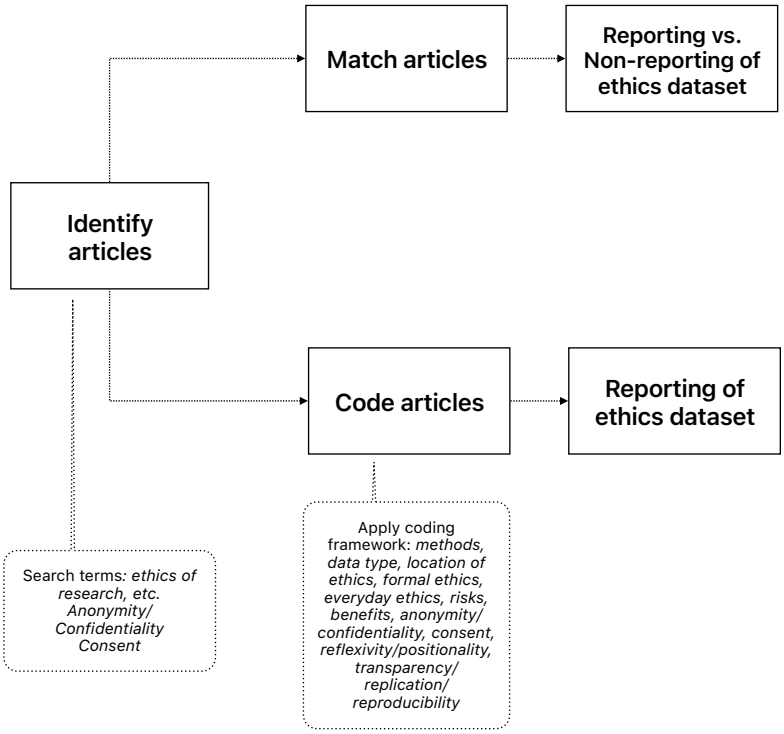


Figure S1. The Process of Constructing the Two Datasets

Below, we list the search terms used to identify articles that report ethics (Table S1). In detailing the search terms we used, we do so without reporting the wildcards (e.g., "*") used in searches. We used such wildcards extensively and consistently to ensure the search was as comprehensive as possible.

Originally, we began with a set of fairly specific and simple search terms pertaining to ethics (e.g. "ethics of research", "research ethics"). As described in the main article, following the anonymous reviewers' suggestions (for which we are extremely grateful), we increased the number of search terms to widen the net of possible articles reporting ethics that we could code (e.g. anonymity, consent). Enlarging the search terms, we increased the size of the reporting vs non-reporting dataset from 69 to 237 articles and enlarged the reporting dataset from 243 to 709 articles.

In enlarging the reporting dataset, and the search terms, it is also important to comment on some of the challenges of doing so. In Table S1 we segment search terms by those that are 'simple' and those more complex. By simple search terms, we mean that these terms easily identify articles that discuss, report, and pertain to our interest of reporting ethics.

However, there are a number of more complex terms. These terms, variously, may refer to concepts more commonly used in political science (e.g. whether states 'consent' to certain practices, or actors engage deceive/engage in deception). In these instances, it is necessary to ensure that the use of the term pertains to a research design decision and falls within the realm of discussing and reporting ethics.

A number of other search terms, namely anonymous/anonymity and debrief, are also complex not because they are commonly used political science concepts. Rather, authors can refer to them without necessarily suggesting they

Table S1. Search Terms

Simple search terms	More complex search terms
ethics of research	1. consent
research ethics	2. deceive
ethics review	3. deception
ethics procedure	4. anonymous/anonymity
ethics clearance	5. debrief
ethics board	
ethics committee	
ethics guidelines	
human subjects committee	
human subjects review	
ethics committee	
IRB	
internal review board	
institutional review board	
confidential	

did provide anonymity or debriefed for ethical reasons. For example, authors may refer to anonymity for reasons of data collection (and often rigor of data collection), such as when using anonymous games, rather than to detail explicit questions of ethics. Similarly, authors may refer to "Anonymous interviewee" without reporting why or how they offered or discussed anonymity with interviewees. In both cases, we do not include such articles. We include only articles that detail how or why they provided anonymity or confidentiality, e.g. for participant protection, and a a type of considered practice of research ethics.

Equally, many authors refer to debriefing but do not detail what debriefing entailed. We also do not include such articles unless they detail that debriefing was designed to be a more ethical practice or was designed as a post-deception practice. Finally, as reported in the main article, we also excluded articles that only reported on issues of deception/deceit since this issue falls outside the scope of our coding framework. Our coding framework was designed to be plural with regard to methodology and method, rather than method-specific (with deception being more specific to experimental methods than to other methods).

2 Journals' Editorial, Professional Organization, and Publisher Portals' Ethics Policies

Since we originally began working on this project (and collecting data from 2000-2018), the publication of American Political Science Association (2020) principles has encouraged several journals to adapt and require sharing of institutional ethics protocols at a minimum. We detail these minimum standards in Table S2 that require reporting of institutional ethics approval (e.g. IRB; updated in April 2024). But, the exact content and the nature of what is requested in terms of reporting differs across journals.

Table S2. Journals' Word Count and Ethics Specifications

Journal	Maximum Word Count	Reporting of Ethics Approval Required
AJPS	10,000	Y
APSR	12,000	Y
BJPIR	8,000	
BJPS	10,000	
CP	12,000	
CPS	12,000	Y
IO	14,000	
IPS	10,000	
ISQ	12,000	Only in relation to Pre-analysis Plan
JCR	11,000	
JoP	35 pages	Y
JPART	12,000	Y
JPR	10,000	Only in relation to data transparency
PA	7,250	Y
PB	9,000	Y
PC	10,000	Y
PolS	10,000	
PoP	12,000	Y
PRQ	10,000	
PS	4,000	Y
SD	10,000	Y (if not possible to disclose interview identities)
TWQ	8,000	
WP	12,500	Y

Besides APSA, we checked professional organizations across different relevant national contexts as detailed in Table S3. In addition, we checked the following publishers' submission portals: Cambridge Core, Taylor and Francis, and SAGE.

Table S3. National Political Science Organizations Checked for Research Ethics Policies Including on Reporting

Name of Association	Country or Region
Australian Political Science Association	Australia
Austrian Political Science Association	Austria
Francophone Belgian Association of Political Science	Belgium
Canadian Political Science Association	Canada
European Political Science Association	Europe
European International Studies Association	Europe
French Association of Political Science	France
German Political Science Association	Germany
Indian Political Science Association	India
Italian Political Science Association)	Italy
Italian Association for International Organization	Italy
Latin American Association of Political Science	Latin America
The Association of Political Science of Luxembourg	Luxembourg
Spanish Association of Political Science and Public Administration	Spain
Swiss Political Science Association	Switzerland
British International Studies Association	UK
Political Science Association	UK
International Studies Association	US

3 Number of Articles in the Datasets

Table S4. Number of Articles in the Reporting vs Non-Reporting Ethics and Reporting of Ethics Datasets by Journal (2000-2018)

Journal	Dataset	
	Reporting	***Reporting vs Non-Reporting***
JoP	84 (11.8%)	1, 158
AJPS	49 (6.9%)	996
CPS	20 (2.8%)	949
APSR	32 (4.5%)	562
JCR	24 (3.4%)	521
CP	6 (0.8%)	410
IO	4 (0.6%)	409
PoP	15 (2.1%)	338
WP	14 (2.0%)	286
BJPIR	11 (1.6%)	
BJPS	8 (1.1%)	
EJPR	6 (0.8%)	
IPS	6 (0.8%)	
IS	7 (1.0%)	
ISQ	12 (1.7%)	
JCMS	38 (5.4%)	
JPART	60 (8.5%)	
JPR	18 (2.5%)	
PA	19 (2.7%)	
PB	44 (6.2%)	
PC	11 (1.6%)	
PolS	19 (2.7%)	
PRQ	31 (4.4%)	
PS	122 (17.2%)	
SD	13 (1.8%)	
TWQ	36 (5.1%)	

4 Coding of Mixed-Methods Articles

We code as mixed-methods an article that combines elements of quantitative and qualitative analysis in a substantive way. Both quantitative and qualitative elements must be used or leveraged to test the hypothesis or to further interpret the evidence, through different types of method mixing, following Creswell and Plano Clark (2018). For example, there are mixed-methods articles that use qualitative research in a post hoc manner, to establish the mechanisms for the findings; other articles, however, integrate both quantitative and qualitative analysis from the start. We code, for example, articles that combine the use of quantitative analysis with substantive analysis of case studies and/or case study comparisons, analysis of qualitative data (e.g. interviews or ethnography), or use of QCA or fuzzy sets. Conversely, we do not code as mixed-methods, for example, where a case study is used descriptively as purely a motivation for the article or to derive a hypothesis, or where qualitative data or evidence (e.g. case studies, such as an analysis of voting patterns in two cases, or interviews which are commonly part of a survey) are analysed quantitatively. In both examples, these articles would be coded as quantitative.

The proportion of mixed-methods articles is surprisingly low (2.4% across the Reporting vs Non-Reporting Ethics dataset) but also remains static over time (Figure S2). In comparison, what is more stark is the increase in quantitative methods and the decrease in qualitative methods.

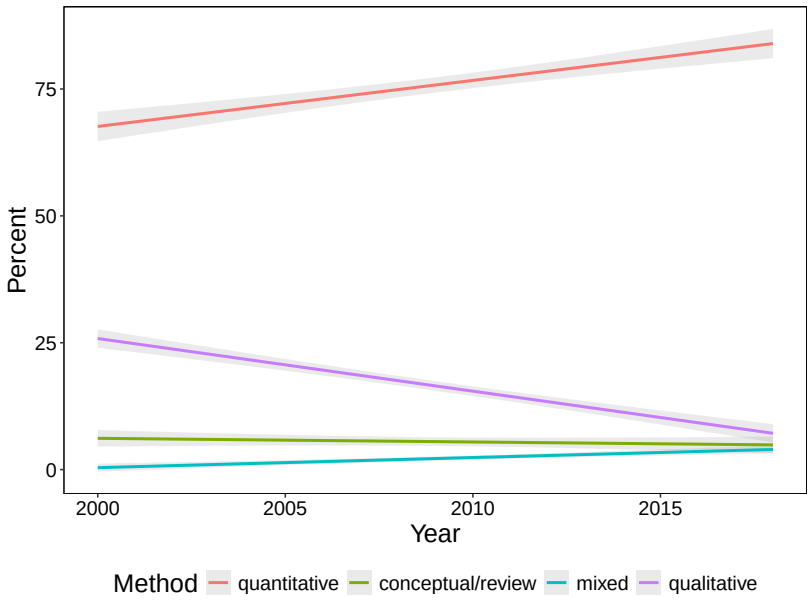


Figure S2. Articles by Method over Time in Reporting vs Non-Reporting Dataset (Smoothed Line)

5 Determining a Baseline Expectation for Ethics Reporting

To determine a baseline expectation for how many articles could or should be reporting ethics, we took a random subset of articles from the reporting vs. non-reporting ethics dataset that did *not* report ethics (560 articles, ≈ 10% of all articles). We take a subset that is representative of the dataset, by stratifying by year, to avoid having to re-code the entirety of the 5,629 articles. In compiling this random subset, we included only quantitative, qualitative, and mixed-methods articles as articles that are empirical and likely collected data.[1]

From this subset, we coded for the methods of data collection used by applying the same coding framework as for the reporting of ethics dataset (Table S5). We consider that those using methods of a field experiment, survey experiment, survey, interviews or focus group, ethnography, and social media data, ought to be reporting ethics since they are all collecting primary data directly from human participants; we find that 23.0% of articles from the subset use such methods.

Table S5. Calculating a Baseline from Methods of Data Collection in Subset of Reporting vs Non-Reporting Dataset

Method	Used (%)	Primary Method (%)	Could/Should Report Ethics	Aggregation of Primary Method (%)
Field Experiment	7.84	3.21	Yes	23.00
Survey	13.37	7.31	Yes	
Survey Experiment	N/A	4.63	Yes	
Interview/focus group	10.34	7.31	Yes	
Ethnography	0.53	0.36	Yes	
Social Media	0.18	0.18	Yes	
Text Data	4.99	3.21	Maybe	23.17
Archival Material	24.96	19.96	Maybe	
Existing Data/Datasets	77.18	52.23	Unlikely	53.83
No Data/Literature Review	1.60	1.60	Unlikely	

Note: For 'Used' we excluded survey experiment and measure these under 'Survey'; we also code for any other methods they used.

We consider that collecting and analyzing archival or textual data may, depending on the data, require consideration and thus reporting of ethics; we find that 23.2% of articles from the subset use such methods. As argued by Subotić (2020) and Hoover Green and Cohen (2021) these methods of data collection are not free from considerations of ethics, whether in terms of the rights of the dead in archival material, or the psychological effects for researchers coding textual data about conflict events. In sum, 46.2% of articles in the subset could and/or should be considering questions of ethics, and therefore could and/or should be reporting questions of ethics. Bearing in mind that only 4.2% articles *are* reporting ethics, we can estimate that only about 10% of articles that should be reporting research ethics are actually reporting it (2000-2018).

Finally, we consider those using no data or existing data/datasets as the least likely category that consider issues of ethics, and could or should be reporting ethics; we find that 53.8% articles in the subset use these methods. We consider this as the least likely category because, depending on the specifics of data, questions of ethics might arise – questions that ought to be reported. Moreover, we find that some articles using existing data/data sets *do* report ethics (in both the reporting and reporting vs. non-reporting datasets). Moreover, projects that rely primarily on existing data/datasets, as well as archival and textual data, may also draw secondarily on other primary methods of data collection for which there would be ethical questions (e.g. surveys or interviews, etc.). While we find no articles in the subset that do use existing data/datasets, archival materials, textual data and such primary sources of data, this should not be excluded as a possibility beyond the parameters of this study.

1. I.e. we excluded articles that were conceptual/review, as articles that are not empirical/collect or use data, but offer methodological discussions/reflections and might draw on authors' prior published empirical work.

This baseline should be seen as contextual in terms of the time period and journals analyzed, in particular in terms of the methods that are used more commonly than others. For example, when comparing to the reporting dataset – where we have a wider and more diverse sample of articles published in political science – in turn we see a wider variety of specific methods of data collection being used (Table S6). In the reporting of ethics dataset, we find fewer articles re-using data and more collecting data, whether via experiments, survey experiments, or interviews. Of course, there are likely to be differences between those that report and do not report ethics, stemming from methods; however, we would also argue that specific journals are also more likely to publish some methods versus others (e.g. experimental work in JOP, PA and PB, interview studies in JPART and JCMS).

Table S6. Methods of Primary Data Collection in Reporting Dataset

Primary Method	Used (%)	Could/Should Report Ethics	Aggregation of Primary Method (%)
Field Experiment	24.76	Yes	87.78
Survey	26.22	Yes	
Survey Experiment	12.21	Yes	
Interview/Focus Group	20.68	Yes	
Ethnography	2.93	Yes	
Social Media	0.98	Yes	
Text Data	0.16	Maybe	3.91
Archival Material	3.75	Maybe	
Existing Data/Datasets	8.14	Unlikely	8.3
No Data/Literature Review	0.16	Unlikely	

6 Gender in the Datasets

6.1 Assigning Gender to the Reporting vs Non-Reporting Dataset

While we had data on author gender from Teele and Thelen (2017), for the data that we added (2015-2018), we used the Wais (2006) genderizeR package – consistent with Teele and Thelen (2017) who used it for their original dataset. As we note in the main article, we are mindful of the limitations of a binary approach to gender and of using names to identify gender. However, this approach is the most suitable for processing large amounts of data at the time of writing.

The Wais (2006) genderizeR package in R assigns a probability of likelihood that the name is male or female by reporting data on "proportion female" and "proportion male" (which are directly inverse, see Figure S3). We define gender-ambiguous names as those with a proportion female between 0.3 and 0.7, following Dion and Mitchell (2019). Identifying ambiguous names, we then hand-coded these for gender using institutional profiles/web-pages and personal web-pages. We followed the same process of hand-coding gender in the Reporting of Ethics dataset. Both approaches of using genderizeR and hand-coding gender are not perfect methods of establishing author gender but they are widely used in existing research, and we follow the approaches of this existing research.

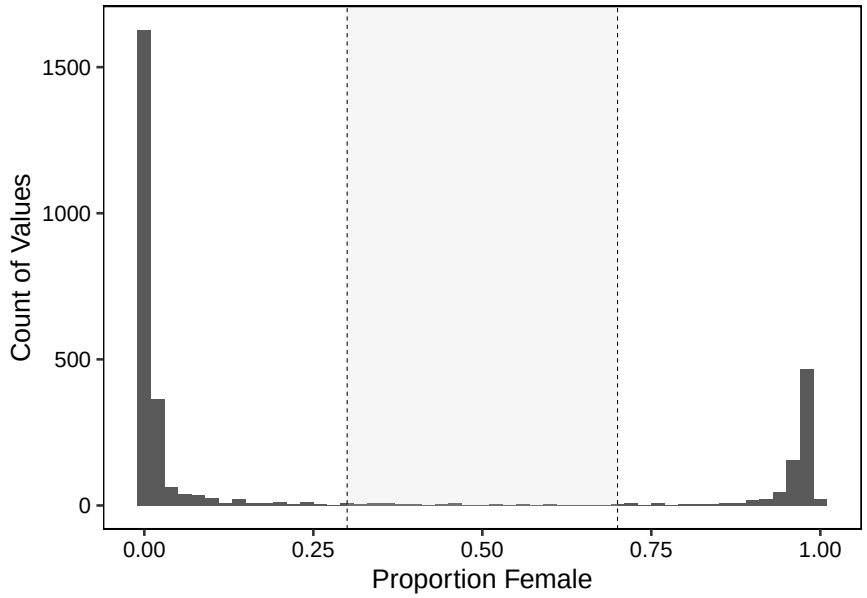


Figure S3. Histogram of Proportion of Women's Names by genderizeR R-package

Note: Shaded area indicates the boundary (0.3 - 0.7) that we hand-coded

In total, we added data for 1,724 articles to the reporting vs. non-reporting dataset which included names of 3,153 authors. Of these, 132 were marked as ambiguous (4.19%). This degree of error/ambiguity is slightly higher than others doing similar (as we report in the main article). For example, Teele and Thelen (2017, note 11) compared the genderizeR package with their hand-coding of gender and found that it was 98% accurate.

7 Regression Analysis

7.1 Choosing a Model

Given discussions over the usefulness of binomial models, and the weight of the binomial dependent variable for the data we are using (95.8 vs 4.2% of observations), one crucial choice was that of model type – between binomial logistic regression, a simple linear regression (Table S7) and Rare Events Logistic Regression (provided by the ReLogit package in R, see Tomz, King, Zeng, *et al.* 2003; Choirat *et al.* 2017; Imai, King, and Lau 2008 and Table S8). It is worth mentioning that the ReLogit and linear models demonstrate the same variables as significant as the binomial model (see Table 3 in main article). Still, we offer some further reasons below for our choice of a binomial model over alternatives.

Table S7. Linear Models

	Dependent Variable: Reporting vs. Non-Reporting of Research Ethics					
	Percent Women			**Categorical Gender Variable**		
	Model 1	**Model 2**	**Model 3**	**Model 4**	**Model 5**	**Model 6**
Percent Women	0.000**		0.000**			
	(0.000)		(0.000)			
Women only (100%)				0.021**		0.022**
				(0.008)		(0.008)
Mixed-gender (1-99%)				0.016*		0.016*
				(0.007)		(0.007)
Method: Conceptual/review	0.006	0.006		0.007	0.006	
	(0.012)	(0.012)		(0.012)	(0.012)	
Method: Mixed	0.105***	0.107***		0.106***	0.107***	
	(0.017)	(0.017)		(0.017)	(0.017)	
Method: Qualitative	-0.010	-0.008		-0.010	-0.008	
	(0.008)	(0.008)		(0.008)	(0.008)	
Journal Year	0.005***	0.006***	0.006***	0.005***	0.006***	0.006***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Author Number	0.007**	0.007*	0.008**	0.006	0.007*	0.006
	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)
(Intercept)	-10.946***	-11.195***	-11.589***	-10.943***	-11.195***	-11.586***
	(0.990)	(0.988)	(0.981)	(0.990)	(0.988)	(0.980)
No. of Observations	5629	5629	5629	5629	5629	5629
R^2	0.037	0.035	0.030	0.037	0.035	0.030
R^2 Adjusted	0.036	0.034	0.029	0.036	0.034	0.029
AIC	-2292.4	-2284.4	-2257.4	-2291.9	-2284.4	-2256.8
BIC	-2239.3	-2238.0	-2224.2	-2232.2	-2238.0	-2217.0
Log Likelihood	1154.182	1149.217	1133.700	1154.957	1149.217	1134.413
F-statistic	35.629	40.704	57.182	30.763	40.704	43.246
Root Mean Square Error	0.20	0.20	0.20	0.20	0.20	0.20

***p<.001; **p<.01; *p<.05

Note: Coefficients with standard errors in parentheses; dependent variable is reporting vs. not-reporting of ethics; the number of observations is 5,629 for all models.

First, we decided for the binomial model over the linear model (Table S7) given the coefficients are more intelligi-

ble for the binomial model. Second, when comparing Akaike Information Criterion (AIC) and Bayesian Information Criterion (BIC) figures for both, the binomial model performs better. In terms of the binomial model over the ReLogit model (Table S8), there is less to decide between since the AIC and BIC figures are identical, the coefficients – and their statistical significance – are extremely similar. For ease of interpretation, accessibility and recognizability, we opted for the binomial model since there is not much indication that we need a model with more sophisticated modeling capacity. Finally, the fact that we get similar results across binomial, linear, and ReLogit models, offers more evidence for the findings from the binomial models.

Table S8. Relogit Models

	Dependent Variable: Reporting vs. Non-Reporting of Research Ethics					
	Percent Women			**Categorical Gender Variable**		
	Model 1	**Model 2**	**Model 3**	**Model 4**	**Model 5**	**Model 6**
Percent women	0.01**		0.01**			
	(0.00)		(0.00)			
Women-only (100%)				0.55**		0.55**
				(0.19)		(0.18)
Mixed-gender (1-99%)				0.41*		0.39*
				(0.17)		(0.17)
Method: Conceptual/review	0.17	0.18		0.18	0.18	
	(0.30)	(0.30)		(0.30)	(0.30)	
Method: Mixed	1.22***	1.27***		1.23***	1.27***	
	(0.25)	(0.25)		(0.26)	(0.25)	
Method: Qualitative	−0.50	−0.42		−0.50	−0.42	
	(0.28)	(0.28)		(0.28)	(0.28)	
Journal year	0.18***	0.19***	0.19***	0.18***	0.19***	0.19***
	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)
Author number	0.17**	0.15*	0.17**	0.13	0.15*	0.14*
	(0.06)	(0.06)	(0.06)	(0.07)	(0.06)	(0.07)
(Intercept)	−370.20***	−376.44***	−386.72***	−370.53***	−376.44***	−386.95***
	(35.73)	(35.66)	(35.44)	(35.75)	(35.66)	(35.46)
No. of Observations	5629	5629	5629	5629	5629	5629
AIC	1765.21	1773.28	1783.49	1765.52	1773.28	1783.92
BIC	1811.66	1813.10	1810.03	1818.61	1813.10	1817.10
Log Likelihood	−875.61	−880.64	−887.74	−874.76	−880.64	−886.96
Deviance	1751.21	1761.28	1775.49	1749.52	1761.28	1773.92

*** p<.001; ** p<.01; * p<.05

Note: Coefficients with standard errors in parentheses; dependent variable is reporting vs. not-reporting of ethics; the number of observations is 5,629 for all models.

7.2 First-Author Gender

For robustness, we also check whether the insights from the binomial regression analysis also hold when considering a first-author's gender (Table S9). However, it must be mentioned, that political science does not have a strong or consistent norm regarding whether a first-author is lead author or whether authors are listed alphabetically.

We run the same binomial model as the original regression analysis, with the dependent variable as reporting vs non-reporting, and again use journal year and author number as controls (Table S9). Controlling for all other variables, the odds of reporting ethics are multiplied by a factor 1.594 of when the first author is a woman ($p<0.01$, model 1), i.e. they are 59.4% higher. We therefore find a similar gendered effect to the original models when analyzing the gender of co-authors.

Table S9. Binomial Models with First-Author's Gender

	Dependent Variable: Reporting vs. Non-Reporting of Research Ethics			
			Categorical Gender Variable	**Percent Gender Variable**
	Model 1	**Model 2**	**Model 3**	**Model 4**
First-author: Woman	1.594**	1.580**	1.220	1.296
	[1.199 - 2.107]	[1.192 - 2.082]	[0.756 - 1.969]	[0.778 - 2.154]
Women only (100%)			1.415	
			[0.781 - 2.549]	
Mixed-gender (1-99%)			1.366	
			[0.899 - 2.038]	
Percent Women				1.003
				[0.997 - 1.009]
Method: Conceptual/review	1.152		1.157	1.149
	[0.619 - 2.003]		[0.612 - 2.012]	[0.608 - 1.997]
Method: Mixed	3.385***		3.378***	3.354***
	[2.021 - 5.477]		[2.004 - 5.477]	[1.992 - 5.430]
Method : Qualitative	0.594		0.590	0.587
	[0.330 - 0.995]		[0.327 - 0.989]	[0.326 - 0.984]
Journal Year	1.200***	1.210***	1.201***	1.200***
	[1.161 - 1.244]	[1.170 - 1.254]	[1.161 - 1.244]	[1.160 - 1.244]
Author Number	1.174*	1.176**	1.145	1.181**
	[1.033 - 1.326]	[1.037 - 1.323]	[0.988 - 1.313]	[1.039 - 1.334]
(Intercept)	0.000***	0.000***	0.000***	0.000***
	[0-0]	[0-0]	[0-0]	[0 - 0]
No. of Observations	5629	5629	5629	5629
AIC	1765.1	1783.6	1766.9	1766.2
BIC	1811.6	1810.1	1826.6	1819.3
Log Likelihood	-875.568	-887.783	-874.428	-875.108
F-statistic	27.870	47.070	21.050	23.934
Root Mean Square Error	0.20	0.20	0.20	0.20

*** $p<.001$; ** $p<.01$; * $p<.05$

Note: Odds ratios with 95% confidence intervals in parentheses; dependent variable is reporting vs. not-reporting of ethics; the number of observations is 5,629 for all models.

Similar to the original binomial models, we see the same effect of journal year, author number, and methods. Increasing the publication year by one year, the odds of reporting ethics are multiplied by a factor of 1.200 ($p<.001$) controlling for all other variables, i.e. they are 20.0% higher, reinforcing our finding that ethics reporting is increasing over time. Similarly, for methods, we find a similar effect to the original model without first authors. For mixed-methods articles, the odds of reporting ethics are 3.385 times the odds for quantitative articles controlling for all

other variables ($p<0.001$), i.e. they are 238.5% higher. The odds of ethics reporting in conceptual/review articles or qualitative articles are not significantly higher or lower compared to the odds for quantitative articles. Finally, as per the original model, every additional author multiplies the odds of ethics reporting by 1.174 ($p<0.05$), controlling for all other variables, i.e. increases them by 17.4%.

However, when first-author gender and other author gender variables are included (whether percent women/model 4 or the categorical gender variable/model 3), the effect of gender is cancelled out. This effect is anticipated since these variables replicate each other, especially in instances of single authorship (38.8% of articles).

7.3 Potential Co-Correlation Between Variables

We might anticipate some potential co-correlation between some of the variables used in the regression models, namely between the number of authors and the proportion of women authors, between methods and the proportion of women authors, and between methods and the number of authors. For example, we might anticipate quantitative work to be published by larger authoring teams versus qualitative work. These potential correlations might affect reporting of ethics, and the findings that we present in this article.

In fact, we find relatively little correlation between these variables, as indicated in Table S10. For example, while the proportion of women authors is negatively correlated with author number, which is to say that the number of women decreases as the number of authors increases, this correlation is low (-0.07). For other variables, converting categorical data to numeric, we also observe little correlation between the proportion of women and method (0.11), and method and author number (-0.26). To that end, the most correlated are method and author number which is to say that quantitative articles are slightly likely to have more authors than other methods.

Put together, the limited correlation between these variables provides further evidence for our findings that women, and more women, are more likely to lead to ethical reporting vs men authors.

Table S10. Investigating Potential Correlation between Variables

	Percent Women	**Author Number**	**Method**
Percent Women	/		
Author Number	−0.07	/	
Method	0.11	−0.26	/

8 Reporting Dataset

8.1 Elaboration of Coding Framework

8.1.1 Coding Examples: Everyday Ethics

We distinguish in our coding framework between formal/institutional ethics (e.g. IRB) and what Guillemin and Gillam (2004) describe as "everyday ethics" or "ethics in practice". Here, we code articles that discuss or report on issues or challenges of ethics that emerged in the process of collecting and/or analyzing data. These might, for example, arise in the process of conducting research in an unexpected, unanticipated, or prescient way, typically beyond how research was designed or anticipated in advance, and in advance of preparing for formal/institutional ethics review.

In the example below, Denov (2006, 324) notes the emerging ethical challenges in the course of collecting data, namely of power asymmetries between participants and researchers:

> *The research posed other important ethical challenges, particularly with regard to the profound disparities of power existing between adult researchers and child participants. The research team was thus highly aware of the potential for re-victimization, making the ethical implications of the research a central aspect of the ongoing training and dialogue of the entire research team. Support structures were put into place to ensure that child participants were provided with ongoing support and assistance during and in the aftermath of interviews.*

While there is also a discussion of other aspects, e.g. risks to participants via re-traumatization, we are also interested to code the emergent and prescient challenges as described by Denov (2006, 324).

In another example, Muldoon *et al.* (2008, 693) describe their use of an innovative method of data collecting by asking participants to write letters, as a way to overcome some of the ethical (and other challenges) of seeking out this kind of information. They also note the limitations of taking this more ethical approach:

> *this study has offered insights into young people's perceptions and appraisal of paramilitary groups, as well as the social and psychological motivations that they perceive may underlie involvement with these groups. While the essay approach adopted surmounted many of the practical, ethical and methodological difficulties associated with researching this sensitive topic with young people, it is, nonetheless, constrained by a number of limitations. […] The explanations offered were those of young people who may or may not have engaged in political violence previously. Owing to the sensitive and ethical issues associated with asking such questions, we have no access to young people's actual experience of political violence at a personal or family level as either victims or perpetrators.*

8.2 Author Rank

To investigate further trends in terms of who reports ethics, we also collected and coded data for author rank.[2] With 91.26% of articles having three or fewer authors, we coded only for the first three authors.

For some journals, coding an author rank at the point of article publication was facilitated by it being included in articles, e.g. in author biographies (APSR and PS, although not for the entire time period back to 2000).

However, most journals in the dataset over the time period do not include information on author rank. Instead, we sought out author CVs to establish their rank at the point of publication. In most instances, CVs were available; in cases where they were not, personal and institutional web-pages or other articles/publications were sought out from the period that the article was published to establish author rank at the point of publication. This method of checking CVs is likely to incur some error since the coder must decide, for example, if someone is promoted within the year that an article is published for which position to code. Moreover, some judgment calls needed to be made when an individual had a dual affiliation, or a transitional affiliation (e.g. from PhD student to Assistant Professor). We followed the author's institution reported in the article and from here took their rank at the point of publication.

Moreover, there are also differences in rank across different higher education contexts. Given that most authors in the dataset are from US-based institutions, and that US ranks are increasingly present elsewhere (e.g. in some UK universities), we coded according to US rankings: student, in a PhD program, Post Doc, Assistant Professor, Associate Professor, Full Professor, independent author (i.e. not affiliated with a university) and Fellow, where an individual might be employed in a research-only position that does not align with any of the other categories.

That being said, we could locate data for 99.3% of the authors; for these authors, we report their author rank as "unknown" (0.7% of authors).

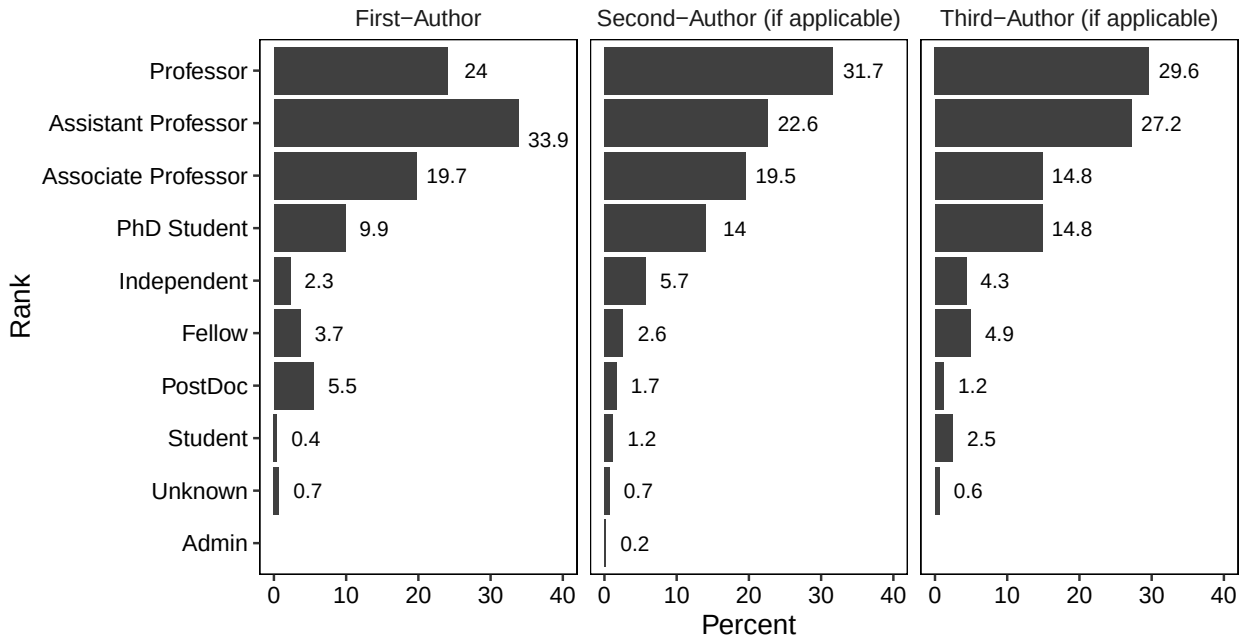


Figure S4. Author Rank

2. We thank the anonymous reviewers for encouraging us to investigate author rank for those reporting ethics.

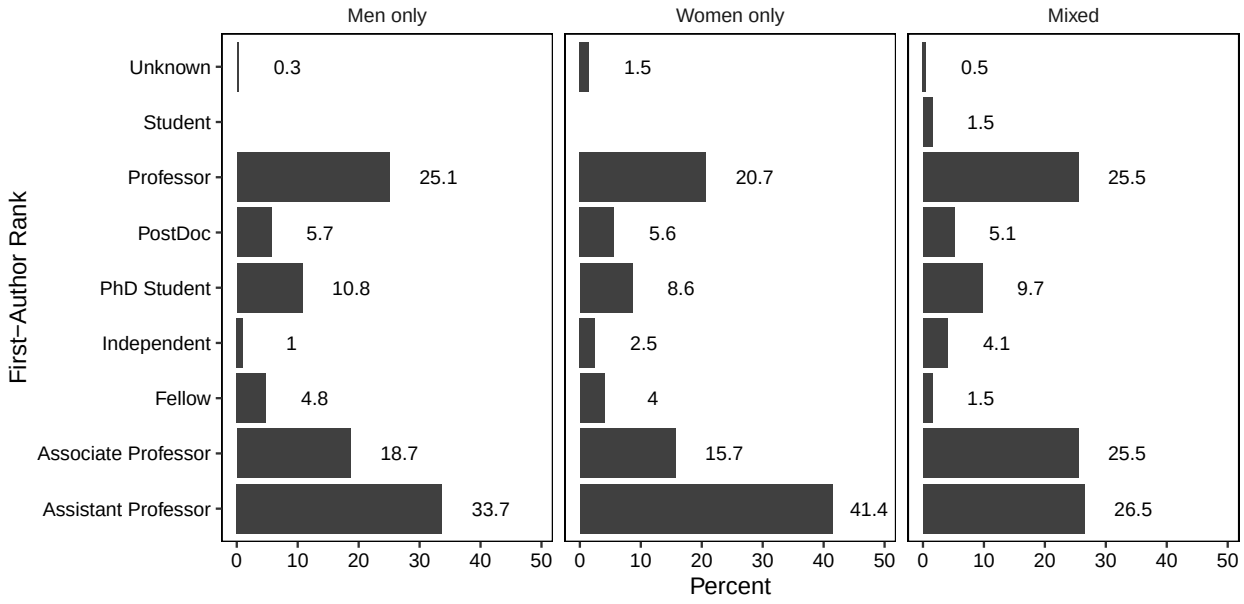


Figure S5. Rank of First-Authors by Gender

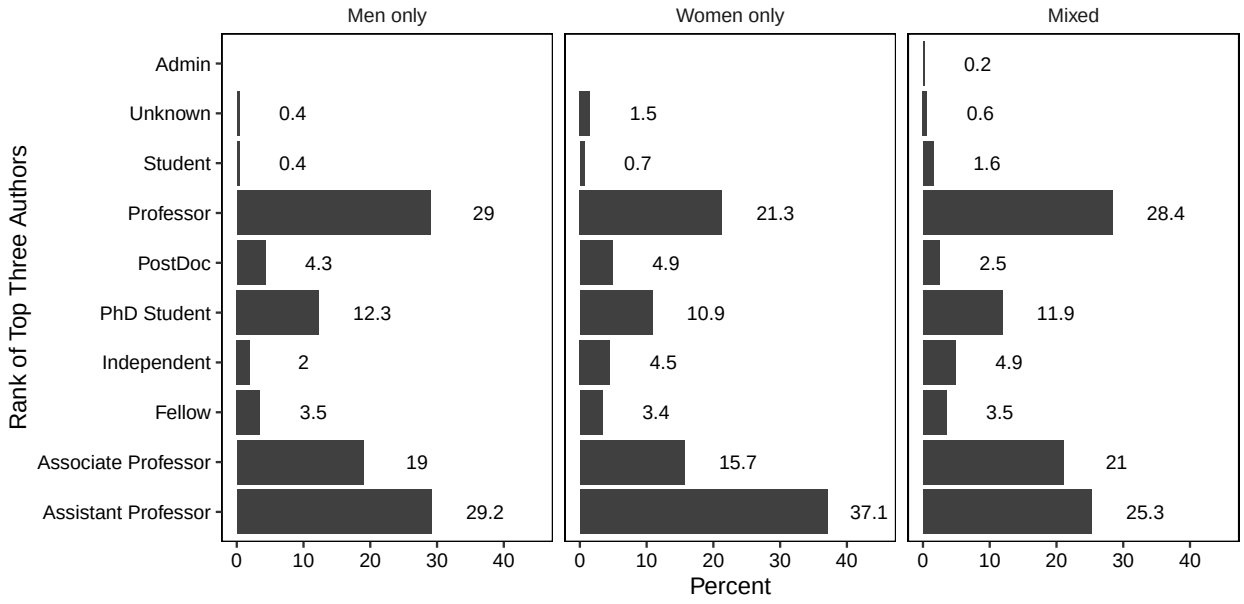


Figure S6. Rank of Top-Three Authors by Gender

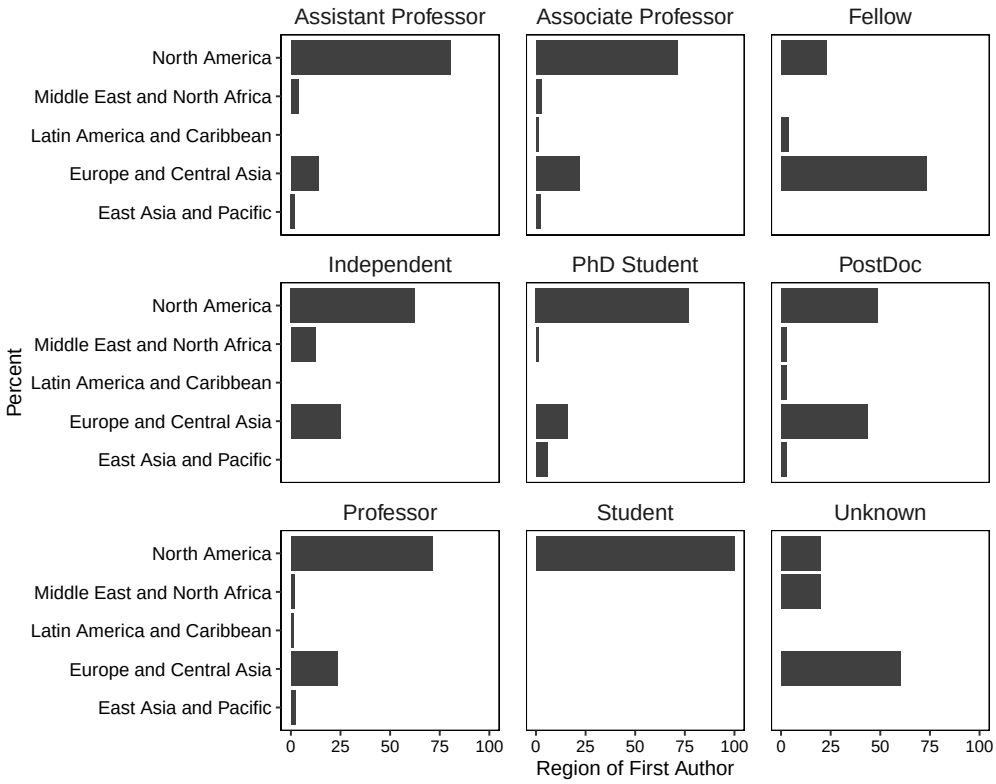
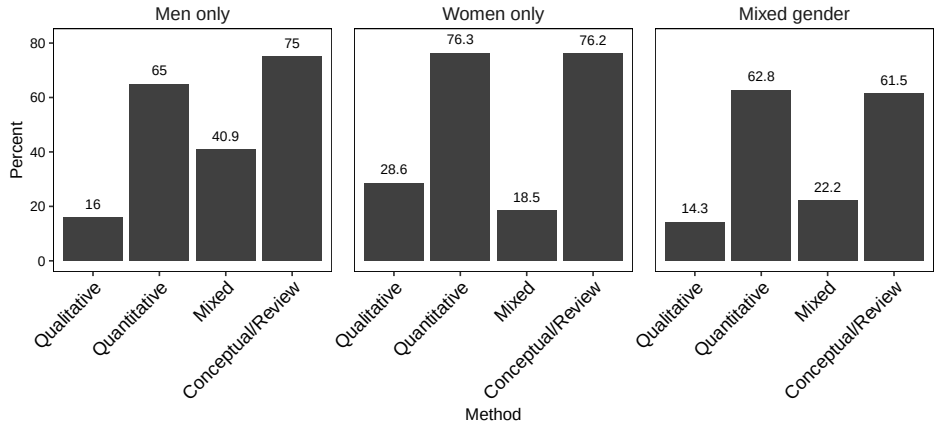


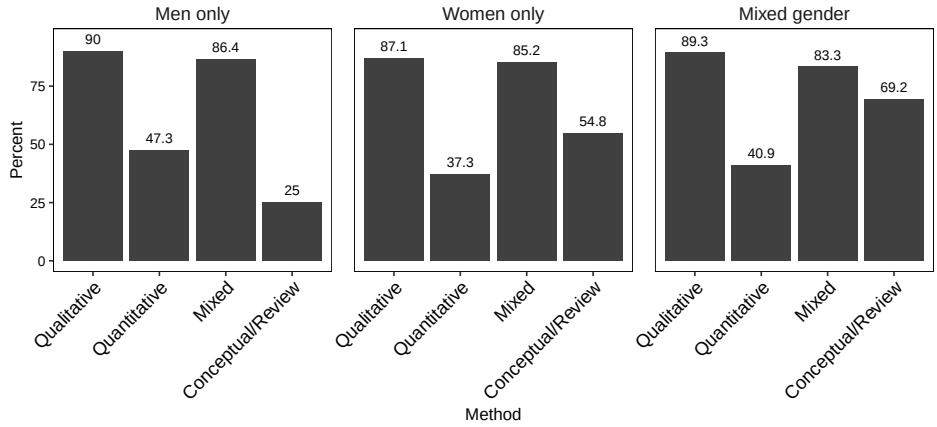
Figure S7. First Author Rank by Region

8.3 Additional Graphs

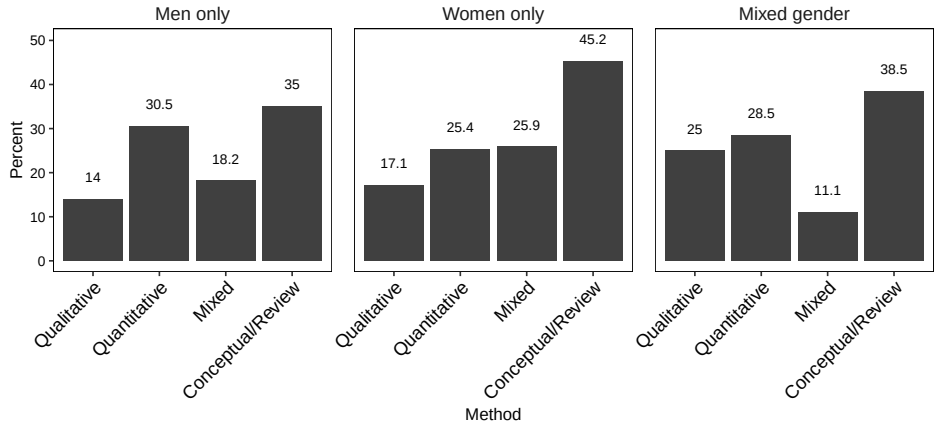
8.3.1 Intersecting Gender and Method for Reporting of Ethics



(a) Percent Reporting of Formal Ethics by Gender and Method

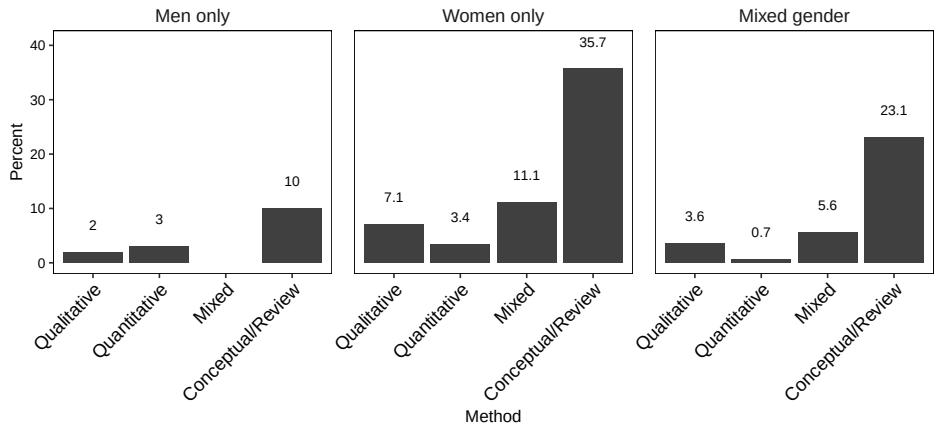


(b) Percent Reporting of Confidentiality and Anonymity by Gender and Method



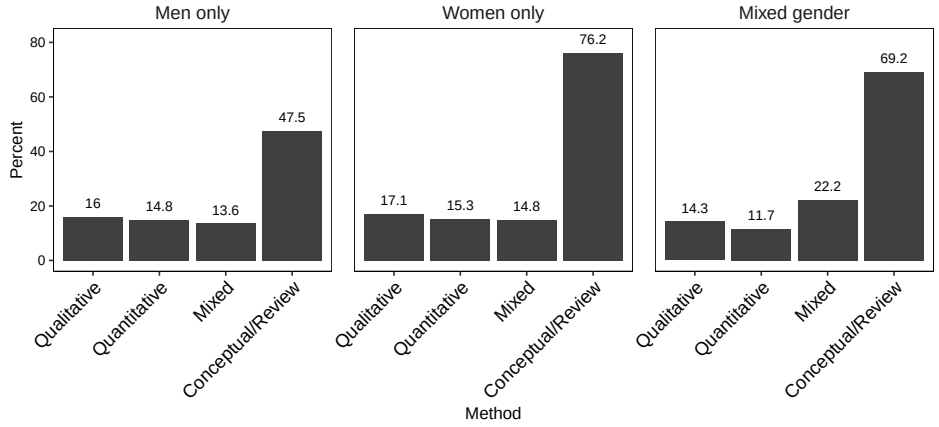
(c) Percent Reporting Consent Protocols by Gender and Method

Figure S8. Dimensions of Ethics by Gender and Method – Percent Reporting Formal Ethics, Confidentiality & Anonymity, Consent Protocols

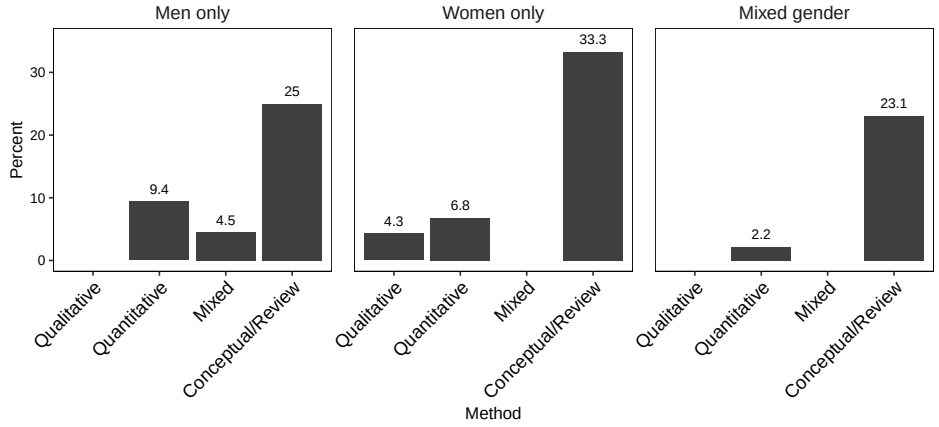


(a) Percent Reporting Everyday Ethics by Gender and Method

Note: Empty categories = 0%



(b) Percent Reporting Risks by Gender and Method



(c) Percent Reporting Benefits by Gender and Method

Note: Empty categories = 0%

Figure S9. Dimensions of Ethics by Gender and Method – Reporting of Everyday Ethics, Risks, and Benefits

8.3.2 Loci of Risks and Benefits

Gender and Loci of Risks and Benefits Intersecting reporting of loci of risks with gender, we see more women reporting risks to participants (25.8%) compared to men (16.5%) and mixed authors (13.8%). We also see more reporting by women of risks to assistants (3.5%) compared to men (assistants: 1.3%) and mixed-gender authors (1.5%). Similarly, women-only authors report more risks to researchers (6.6%), with fewer men-only (4.1%) and mixed-gender co-authors (3.1%) reporting risks to researchers. Thus, across reporting of risks by loci, we observe more reporting by primarily by women and then men compared to mixed-gender authors.

We observe some gender differences concerning reporting of benefits. Women authors report benefits vis-a-vis participants (7.6%), researchers (3%), and assistants (0.5%), more than men (participants: 6.3%, researchers: 1.3%, assistants 0%) and mixed authors (participants: 2%, researchers: 0.5%, assistants 0%); whereas men report benefits to society (3.5%) more than women (2.5%) and mixed authors (1%).

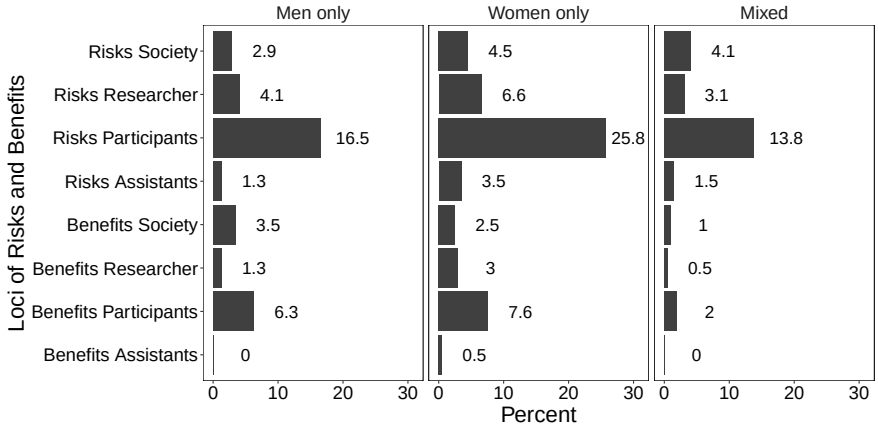


Figure S10. Reporting of Ethics by Loci and Author Gender

Percentage Coded for Presence

Method and Loci of Risks and Benefits We see (like for gender) that risks to participants is the most reported dimension for all methods: conceptual/review (54.7%), mixed (14.9%), qualitative (14.2%), and quantitative articles (11.8%, Figure S11); in other words, quantitative articles appear a slight laggard in reporting or discussing potential risks for participants. This might be because quantitative methods are less risky, or perceived as less risk, to participants. However, without adequate reporting of ethical concerns, it is hard to draw any definitive conclusions. Rather, we would argue that there is no reason that quantitative methods are, inherently, less risky to participants than other methods (especially given concerns raised vis-a-vis methods like field experiments see Teele 2014; Desposato 2018; Michelson 2016), in particular, compared to mixed-method articles that *do* report more risks to participants.

There is even less diversity in reporting benefits, with quantitative articles reporting benefits minimally and only vis-a-vis participants (5%) and society (1.8%), and qualitative articles reporting vis-a-vis participants (2%), researchers (0.7%), and society (0.7%), and mixed-methods articles only reporting benefits to participants (1.5%). In contrast, conceptual/review articles report benefits vis-a-vis participants (15.8%), researchers (9.5%), society (10.5%), and assistants (1.1%, Figure S11).

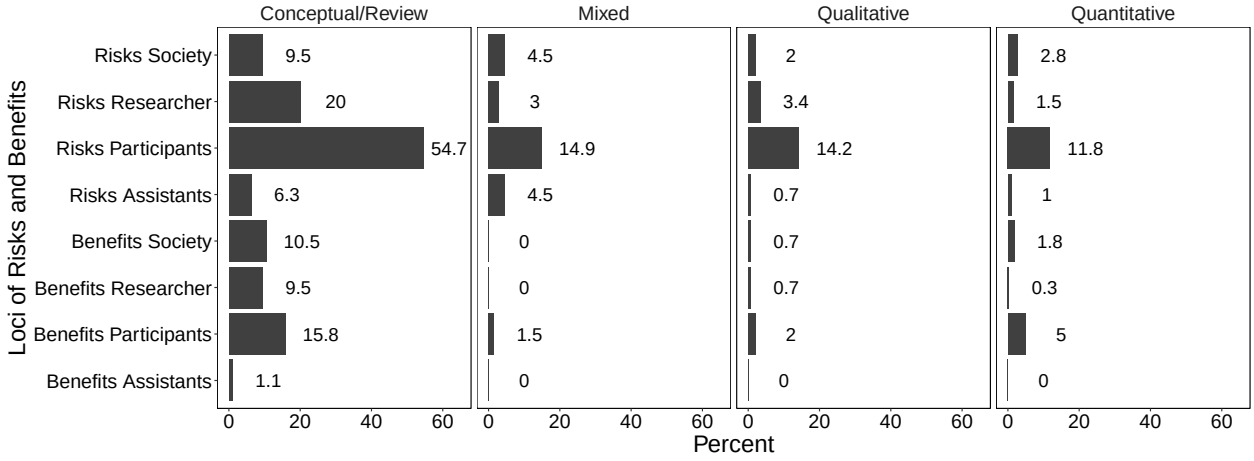


Figure S11. Reporting of Ethics by Loci and Article Method

Note: Percentage Coded for Presence

References

American Political Science Association. 2020. *Principles and Guidance for Human Subjects Research.* Accessed June 29, 2021. https://connect.apsanet.org/hsr/principles-and-guidance/.

Choirat, C, J Honaker, K Imai, G King, and O Lau. 2017. *Zelig: Everyone's Statistical Software, version 5.1. 4.90000.*

Creswell, John W., and Vicki L. Plano Clark. 2018. *Designing and Conducting Mixed Methods Research.* Third Edition. Los Angeles: SAGE.

Denov, Myriam S. 2006. "Wartime Sexual Violence: Assessing a Human Security Response to War-Affected Girls in Sierra Leone." *Security Dialogue* 37 (3): 319–342. https://doi.org/10.1177/0967010606069178.

Desposato, Scott. 2018. "Subjects and Scholars' Views on the Ethics of Political Science Field Experiments." *Perspectives on Politics* 16 (3): 739–750.

Dion, Michelle L., and Sara McLaughlin Mitchell. 2019. "How Many Citations to Women Is "Enough"? Estimates of Gender Representation in Political Science." *PS: Political Science & Politics* 53 (1): 1–7.

Guillemin, M., and L. Gillam. 2004. "Ethics, Reflexivity, and "Ethically Important Moments" in Research." *Qualitative Inquiry* 10 (2): 261–280.

Hoover Green, Amelia, and Dara Kay Cohen. 2021. "Centering Human Subjects: The Ethics of "Desk Research" on Political Violence." *Journal of Global Security Studies* 6 (2): 1–17.

Imai, Kosuke, Gary King, and Olivia Lau. 2008. "Toward a Common Framework for Statistical Analysis and Development." *Journal of Computational and Graphical Statistics* 17 (4): 892–913.

Michelson, Melissa R. 2016. "The Risk of Over-reliance on the Institutional Review Board: An Approved Project is not Always an Ehical Project." *PS: Political Science & Politics* 49 (2): 299–303.

Muldoon, Orla T., Katrina Mclaughlin, Nathalie Rougier, and Karen Trew. 2008. "Adolescents' Explanations for Paramilitary Involvement." *Journal of Peace Research* 45 (5): 681–695. https://doi.org/10.1177/0022343308094330.

Subotić, Jelena. 2020. "Ethics of Archival Research on Political Violence." *Journal of Peace Research* 58 (3): 342–354.

Teele, Dawn Langan. 2014. *Reflections on the Ethics of Field Experiments.* New Haven: Yale University Press.

Teele, Dawn Langan, and Kathleen Thelen. 2017. "Gender in the Journals: Publication Patterns in Political Science." *PS: Political Science & Politics* 50 (2): 433–447.

Tomz, Michael, Gary King, Langche Zeng, *et al.* 2003. "ReLogit: Rare Events Logistic Regression." *Journal of Statistical Software* 8 (i02).

Wais, Kamil. 2006. "Gender Prediction Methods Based on First Names with genderizeR." *The R Journal* 8 (1): 17–37.