

Supplementary File for

Title: Gender equity in health research publishing in Africa

Authors: Baobeid A, Faghani-Hamadani T, Sauer S, et al.

Sensitivity analysis #1:

In the first sensitivity analysis, we considered the error rates among strata defined by the level of information in the author name (initials or missing first name vs full name) and the probability of correct classification generated by the Namsor algorithm. Due to the high error rate (28%) among authors with initials/low classification probability, we conducted a sensitivity analysis to explore the robustness of our results to changes in the gender classification of author names. In particular, we repeated our original analysis described in the main paper, in which the genders were assigned by Namsor, with an adjusted gender classification variable, based on the error rates computed in the data quality check: among authors with initials or missing first name/low classification probability, we randomly selected and changed the classifications of 61.9% of those originally classified as woman to man, and randomly selected and changed the classifications of 23.5% of those originally classified as man to woman.

Table 1: Author names stratified based on probability of correctness.

	Probability of correct classification	Number of authors sampled	Percentage of names sampled from each strata	Error rate [†]
Initials/missing names of authors	< 0.7	187	5%	Overall: 28% Classified as women: 61.9% Classified as men: 23.5%
	0.7-0.9	11	100%	Overall: 0%
Full names of authors	< 0.7	520	2%	Overall: 8% Classified as women: 9.2% Classified as men: 6.6%
	0.7-0.9	38	1%	Overall: 3% Classified as women: 0% Classified as men: 2.7%
	> 0.9	137	1%	Overall: 0%
Overall		893	2.1%	

[†]The genders of the sampled authors were determined manually by searching their institutional web pages and online social network sites such as LinkedIn and ResearchGate. The number of sampled authors with misclassified gender in each stratum was calculated, and error rates were computed by dividing these by the number of sampled authors in the respective stratum. Names that were misclassified were subsequently corrected in our database.

Table 2: Author gender composition by region of affiliation after adjusting for probability classification error rates.

	Authorship position		
	First author (n=6819)	Last author (n=6819)	Single author (n=281)
Overall	p < 0.001 ¹	p < 0.001 ¹	p=0.06 ¹
<i>Men</i>	3632 (53.3%)	4117 (60.4%)	156 (55.5%)
<i>Women</i>	3187 (46.7%)	2747 (40.3%)	125 (45.5%)
Region of author affiliation⁴	p < 0.001 ²	p = 0.003 ²	p < 0.001 ³
Sub-Saharan Africa	n = 3768	n = 3350	n = 146
<i>Men</i>	2257 (59.9%)	2095 (62.5%)	95 (65.1%)
<i>Women</i>	1511 (40.1%)	1255 (37.5%)	51 (34.9%)
Europe/North America	n = 2206	n = 2764	n = 122
<i>Men</i>	868 (39.3%)	1598 (57.8%)	52 (42.6%)
<i>Women</i>	1338 (60.7%)	1166 (42.2%)	70 (57.4%)
Both Sub-Saharan Africa and Europe/North America	n = 729	n = 543	n = 8
<i>Men</i>	442 (60.6%)	327 (60.2%)	5 (62.5%)
<i>Women</i>	287 (39.4%)	216 (39.8%)	3 (37.5%)

Other	n = 116	n = 162	n = 5
<i>Men</i>	65 (56.0%)	97 (59.9%)	4 (80.0%)
<i>Women</i>	51 (44.0%)	65 (40.1%)	1 (20.0%)

¹p-value associated with two-sided one-sample test of proportions, with the null proportion set to 0.5

²p-value associated with Chi-squared test of association between region of affiliation and gender

³p-value associated with Fisher's exact test of association between region of affiliation and gender

⁴p-values associated with one-sample tests of proportions within each region of affiliation and author position category are not included in this table for brevity

Sensitivity analysis #2:

For the second sensitivity analysis, we considered error rates stratified by author region of affiliation, using the same sample of 893 authors. Table 3 shows these error rates, and Table 4 shows the results, based on the corresponding error adjusted gender classification variable.

Table 3: Author names stratified based on region of affiliation.

	Number of authors sampled	Percentage of names sampled from each strata	Error rate
Sub-Saharan Africa	488	2%	<i>Overall: 12.5%</i> <i>Classified as women: 17.9%</i> <i>Classified as men: 9.1%</i>
Europe/North America	288	2%	<i>Overall: 8%</i> <i>Classified as women: 5.4%</i> <i>Classified as men: 10.8%</i>
Both Sub-Saharan Africa and Europe/North America	104	3.7%	<i>Overall: 8%</i> <i>Classified as women: 8.5%</i> <i>Classified as men: 12.3%</i>
Other	13	1.5%	<i>Overall: 3%</i> <i>Classified as women: 20%</i> <i>Classified as men: 12.5%</i>
Overall	893	2.1%	

Table 4: Author gender composition by region of affiliation after adjusting for region-specific error rates.

	Authorship position		
	First author (n=6819)	Last author (n=6819)	Single author (n=281)
Overall	p < 0.001 ¹	p < 0.001 ¹	p =0.51 ¹
<i>Men</i>	3736 (54.8%)	4072 (59.7%)	146 (52.0%)
<i>Women</i>	3083 (45.3%)	2747 (40.3%)	135 (48.0%)
Region of author affiliation⁴	p < 0.001 ²	p < 0.001 ²	p < 0.001 ³
Sub-Saharan Africa	n = 3768	n = 3350	n = 146
<i>Men</i>	2374 (63.0%)	2171 (64.8%)	92 (63.0%)
<i>Women</i>	1394 (37.0%)	1179 (35.2%)	54 (37.0%)
Europe/North America	n = 2206	n = 2764	n = 122
<i>Men</i>	866 (39.3%)	1505 (54.5%)	47 (38.5%)
<i>Women</i>	1340 (60.7%)	1259 (45.5%)	75 (61.5%)
Both Sub-Saharan Africa and Europe/North America	n = 729	n= 543	n = 8
<i>Men</i>	428 (58.7%)	294 (54.1%)	4 (50.0%)
<i>Women</i>	301 (41.3%)	249 (45.9%)	4 (50.0%)
Other	n = 116	n = 162	n = 5

<i>Men</i>	68 (58.6%)	102 (63.0%)	3 (60.0%)
<i>Women</i>	48 (41.3%)	60 (37.0%)	2 (40.0%)

¹p-value associated with two-sided one-sample test of proportions, with the null proportion set to 0.5

²p-value associated with Chi-squared test of association between region of affiliation and gender

³p-value associated with Fisher's exact test of association between region of affiliation and gender

⁴p-values associated with one-sample tests of proportions within each region of affiliation and author position category are not included in this table for brevity