



Quantifying Sex Bias in Clinical Studies at Scale With Automated Data Extraction

Sergey Feldman, PhD; Waleed Ammar, PhD; Kyle Lo, MS; Elly Trepman, MD; Madeleine van Zuylen, BS; Oren Etzioni, PhD

Abstract

IMPORTANCE Analyses of female representation in clinical studies have been limited in scope and scale.

OBJECTIVE To perform a large-scale analysis of global enrollment sex bias in clinical studies.

DESIGN, SETTING, AND PARTICIPANTS In this cross-sectional study, clinical studies from published articles from PubMed from 1966 to 2018 and records from Aggregate Analysis of ClinicalTrials.gov from 1999 to 2018 were identified. Global disease prevalence was determined for male and female patients in 11 disease categories from the Global Burden of Disease database: cardiovascular, diabetes, digestive, hepatitis (types A, B, C, and E), HIV/AIDS, kidney (chronic), mental, musculoskeletal, neoplasms, neurological, and respiratory (chronic). Machine reading algorithms were developed that extracted sex data from tables in articles and records on December 31, 2018, at an artificial intelligence research institute. Male and female participants in 43 135 articles (792 004 915 participants) and 13 165 records (12 977 103 participants) were included.

MAIN OUTCOMES AND MEASURES Sex bias was defined as the difference between the fraction of female participants in study participants minus prevalence fraction of female participants for each disease category. A total of 1000 bootstrap estimates of sex bias were computed by resampling individual studies with replacement. Sex bias was reported as mean and 95% bootstrap confidence intervals from articles and records in each disease category over time (before or during 1993 to 2018), with studies or participants as the measurement unit.

RESULTS There were 792 004 915 participants, including 390 470 834 female participants (49%), in articles and 12 977 103 participants, including 6 351 619 female participants (49%), in records. With studies as measurement unit, substantial female underrepresentation (sex bias \leq -0.05) was observed in 7 of 11 disease categories, especially HIV/AIDS (mean for articles, -0.17 [95% CI, -0.18 to -0.16]), chronic kidney diseases (mean, -0.17 [95% CI, -0.17 to -0.16]), and cardiovascular diseases (mean, -0.14 [95% CI, -0.14 to -0.13]). Sex bias in articles for all categories combined was unchanged over time with studies as measurement unit (range, -0.15 [95% CI, -0.16 to -0.13] to -0.10 [95% CI, -0.14 to -0.06]), but improved from before or during 1993 (mean, -0.11 [95% CI, -0.16 to -0.05]) to 2014 to 2018 (mean, -0.05 [95% CI, -0.09 to -0.02]) with participants as the measurement unit. Larger study size was associated with greater female representation.

CONCLUSIONS AND RELEVANCE Automated extraction of the number of participants in clinical reports provides an effective alternative to manual analysis of demographic bias. Despite legal and policy initiatives to increase female representation, sex bias against female participants in clinical studies persists. Studies with more participants have greater female representation. Differences

(continued)

Key Points

Question What is the magnitude of female underrepresentation in clinical studies?

Findings In this cross-sectional study, machine reading to extract sex data from 43 135 published articles and 13 165 clinical trial records showed substantial underrepresentation of female participants, with studies as measurement unit, in 7 of 11 disease categories, especially HIV/AIDS, chronic kidney diseases, and cardiovascular diseases. Sex bias in articles for all categories combined was unchanged over time with studies as the measurement unit but improved with participants as measurement unit.

Meaning This study suggests that sex bias against female participants in clinical studies persists, but results differ when studies vs participants are the measurement units.

+ Supplemental content

Author affiliations and article information are listed at the end of this article.

Open Access. This is an open access article distributed under the terms of the CC-BY License.

Abstract (continued)

between sex bias estimates with studies vs participants as measurement unit, and between articles vs records, suggest that sex bias with both measures and data sources should be reported.

JAMA Network Open. 2019;2(7):e196700. doi:10.1001/jamanetworkopen.2019.6700

Introduction

For proper application of clinical study results, enrolled participants should represent the populations for which treatments are intended. When female patients receive treatment based on the results of studies of male participants, unanticipated adverse events may occur because of sex-specific differences in disease patterns, metabolism, and drug pharmacokinetics and clearance.^{1,2} Health risks were greater in female patients than in male patients for 8 of 10 prescription drugs withdrawn from the US market from 1997 to 2000.³ The slower metabolism of the insomnia drug zolpidem in female patients than in male patients may have contributed to multiple zolpidem-related motor vehicle crashes before the recommended dose was decreased in female patients by 50%.⁴⁻⁶ Female patients may experience more adverse drug reactions, more disease and disability, later diagnosis, less aggressive treatment, and lower case survival rates for some diseases than male patients.⁷⁻⁹

The National Institutes of Health Revitalization Act of 1993 established legal requirements and guidelines to ensure the inclusion of female participants and racial/ethnic minority participants in clinical research.¹⁰ However, underrepresentation of female participants in studies relative to disease prevalence (known as *enrollment sex bias* or *sex bias*) persists.^{11,12} In treatment trials of 11 non-sex-specific cancers (9671 patients), underrepresentation of female participants was noted in trials of 3 cancer types.¹³ In 120 randomized clinical trials (total, 160 801 participants) in 12 specialties, 24.6% of participants were female, with no improvement observed in sex-balanced enrollment or sex-specific analyses.¹¹ From 2000 to 2002, female participants had lower enrollment fraction—defined as the number of trial participants divided by the estimated number of cancer cases in the population—than male participants for colorectal (total, 8434 participants) and lung cancer (4297 participants) trials.¹² A literature search for 1999 to 2018 showed 13 major analyses of sex bias in clinical studies, but these analyses were limited in size (range, 36-865 studies and 2339-398 801 participants) and disease categories and were performed with manual methods or analysis of isolated data sets (eAppendix and eTable 1 in the [Supplement](#)).¹²⁻²⁴

Computerized, automated data extraction (also known as *machine reading*) of published research articles enables the development of large, complex systems to organize, integrate, and communicate information from numerous studies.²⁵⁻²⁹ However, a literature search did not show previous studies of machine automation for quantifying sex bias in clinical studies at the national or global scale.

The purpose of this study was to develop a scalable automated machine reading method to extract sex data from numerous clinical studies and analyze sex bias in published articles and clinical trial records at scale.^{30,31} We hypothesized that computerized data extraction from numerous articles and records may provide comprehensive and longitudinal information about sex bias in clinical studies at scale.

Methods

Data Sources

We analyzed the number of male and female participants in clinical studies that were identified and extracted in electronic searches from 2 sources on December 31, 2018: (1) published articles from the search engine Semantic Scholar, which had 41 million articles indexed, including more than 20 million full-text articles and all articles in PubMed Central from 1966 to 2018,^{30,32} and (2) clinical trial records

in the Aggregate Analysis of ClinicalTrials.gov (AACT) database, which contained metadata for 288 515 studies registered at ClinicalTrials.gov in 205 countries from 1999 to 2018.^{33,34}

Global disease prevalence data for male and female participants were obtained from the Global Health Data Exchange (GHDx), a database synthesized from multiple data sources, including scientific literature and population representative surveys.^{35,36} Prevalence values for selected disease categories defined by GHDx were obtained from an online catalog of health-related data (eTable 2 in the [Supplement](#)).³⁵

This study was not considered human subjects research according to the Federal Policy for the Protection of Human Subjects because it was a secondary analysis of data from published articles and trial records. Therefore, the study was not submitted for institutional review board approval.

Study Sample and Data Extraction

We identified all articles related to clinical studies in PubMed using article categories selected from the XML PubMed publication type attribute <PublicationTypeList> (1 038 324 articles) (eTable 3 in the [Supplement](#)).³⁷ Semantic Scholar accessed the full text of 388 227 articles (37%). We restricted the analysis to articles about medical disorders by including only articles labeled with any Medical Subject Headings (MeSH) terms under "disease," "vaccination," "disorder," "pathological," or "neoplasms" in the MeSH taxonomy tree, and processed these articles with optical character recognition (OmniPage; Nuance Communications) (295 139 articles). As the analysis was based on automated extraction of male and female participant numbers from tables, we included articles with at least 1 table extracted (249 845 articles).

We developed an algorithm (PubMed-Extract) to extract articles and sex data from tables of articles in portable document format (eTable 4 in the [Supplement](#)). PubMed-Extract was designed to parse the tables, identify relevant semantics of rows and columns by matching patterns, and aggregate information across table rows and columns (eAppendix in the [Supplement](#)). We limited the analysis to 11 GHDx disease categories for which morbidity frequency data were available in GHDx and more than 1000 articles were identified: cardiovascular diseases, diabetes, digestive diseases, hepatitis (types A, B, C, and E), HIV/AIDS, kidney diseases (chronic), mental disorders, musculoskeletal disorders, neoplasms, neurological disorders, and respiratory diseases (chronic). We mapped articles to disease categories using the MeSH terms associated with each article (eTable 5 in the [Supplement](#)). In the 249 845 articles that were processed by optical character recognition and had at least 1 table extracted, 147 807 articles (59%) were mapped to at least 1 disease category, from which PubMed-Extract extracted male and female participant numbers in 43 135 articles (17%).

We developed another algorithm (AACT-Query) to extract sex data from tables in AACT records that could be queried with Structured Query Language. We identified AACT records of 33 361 studies that had male and female participant numbers. After excluding incomplete studies, there were 28 187 studies. After mapping records to disease categories using MeSH terms, we retained 13 165 records (47%) that mapped to at least 1 disease category, and used AACT-Query to extract male and female participant numbers.

Variables

Female prevalence fraction (F-Prev) for each disease category was defined as the fraction of female participants in the disease category and was estimated by dividing the global morbidity count for female participants by global morbidity count for both male and female participants using GHDx data. Female participant fraction (F-Particip) was defined as the fraction of female participants among all participants who were included in the studies, and was estimated 2 ways: with (1) studies as measurement units, by computing the ratio of female participants to all participants for each study and determining the simple average of this ratio for all studies without any weighting by study size and (2) participants as measurement units, by dividing the total number of female participants in all studies by the total number of male and female participants in all studies combined. The female

participant fraction was estimated from articles using PubMed-Extract and records using AACT-Query. The primary outcome variable was enrollment sex bias in clinical studies, defined as $F\text{-Particip}$ minus $F\text{-Prev}$ (values for sex bias ranged from -1 to 1 , with 0 indicating no bias; negative sex bias indicates that female participants were represented less than male participants).

Accuracy of PubMed-Extract Estimates

We evaluated the accuracy of sex bias estimates from PubMed-Extract by comparing them with the true $F\text{-Particip}$ that was determined from manually extracted numbers of male and female participants from 100 randomly selected articles. Mean absolute error was calculated by averaging the absolute difference between the PubMed-Extract estimates and true value of $F\text{-Particip}$ in individual articles.

We evaluated the recall of PubMed-Extract, defined as the percentage of articles for which PubMed-Extract produced the exact number of male and female participants as manually extracted in another random set of 100 articles on cardiovascular diseases. Mean absolute error was sensitive to severity of estimation errors, whereas recall equally penalized all estimation errors.

Comparison Between PubMed-Extract and AACT-Query

To evaluate differences between sex bias estimated with PubMed-Extract vs AACT-Query, we analyzed studies that were represented in both estimates. We identified 1400 articles for which (1) PubMed-Extract produced numerical estimates of sex bias, (2) the articles were linked each to exactly 1 AACT record, (3) the AACT record included numbers of male and female participants, and (4) the full text of the articles was available through PubMed. We compared the numbers of male and female participants between these articles and records and manually inspected a sample of 50 discordant articles and records to determine the reasons for discrepancies. We contacted study authors for comments when we were unable to determine reasons for discrepancies.

Statistical Analysis

For each disease category, we computed 1000 bootstrap estimates of sex bias by resampling individual studies with replacement. Sex bias was reported as mean and 95% bootstrap confidence interval, determined from the bottom 2.5% and top 97.5% of bootstrap estimate percentiles. The P value for the null hypothesis of zero sex bias was equal to the probability of type I error corresponding to the widest confidence interval that contained zero. We calculated P values under the null hypothesis by repeating the bootstrap confidence interval procedure over a fine grid of confidence levels (decreasing from 99.999%), taking the smallest confidence level whose interval contained zero; the P value was the probability of type I error = $2 \times (1 - \text{confidence level})$. For each disease category and time period, statistical significance for a hypothesis test for sex bias was defined by $P \leq .001$ using 2-tailed tests.

For analysis of sex bias in articles vs time, we fitted an intercept-only linear model to sex bias values before or during 1993 and subsequent 5-year increments separately with studies and participants as measurement unit and plotted estimated intercept coefficients vs time with error bars representing 95% confidence intervals for the mean coefficient. We assumed Gaussian distribution because bootstrapping was precluded by dividing the data into 5-year increments.

The association between estimated sex bias and number of participants in each study was evaluated with fixed-effects linear regression, with number of participants defined as a categorical variable with 10 equal-sized bins (eTable 6 in the [Supplement](#)). We controlled for publication year (continuous variable) and disease category (categorical variable). Analyses were performed with the statistical functions of the Python programming language, version 3.6 (Python Software Foundation).

Results

There were 792 004 915 participants, including 390 470 834 female participants (49%), in articles and 12 977 103 participants, including 6 351 619 female participants (49%) in records. The F-Prev was highest for digestive diseases and lowest for hepatitis (Table). With studies as measurement unit, substantial female underrepresentation (sex bias ≤ -0.05) in articles and records was observed in 7 of 11 disease categories, including HIV/AIDS (mean for articles, -0.17 [95% CI, -0.18 to -0.16]), kidney diseases (chronic) (mean, -0.17 [95% CI, -0.17 to -0.16]), cardiovascular diseases (mean, -0.14 [95% CI, -0.14 to -0.13]), neoplasms, digestive diseases, neurological disorders, and hepatitis (Table). The only category with female overrepresentation was musculoskeletal disorders (Table).

With participants as measurement unit, sex bias against female participants in articles was highest for chronic kidney diseases and lowest for musculoskeletal disorders and HIV/AIDS, and in records was highest for HIV/AIDS, chronic kidney diseases, and cardiovascular diseases. Sex bias usually was less negative when the measurement unit was participants vs studies (eg, for articles

Table. Sex Bias in Clinical Studies Determined From Published Articles and Clinical Trial Records^a

Disease Category	Global Female Prevalence Fraction	Measurement Unit	Published Articles			AACT Records		
			Studies or Participants, No.	Female Participant Fraction	Sex Bias (95% CI)	Studies or Participants, No.	Female Participant Fraction	Sex Bias (95% CI)
Cardiovascular	0.51	Studies	14 371	0.37	-0.14 (-0.14 to -0.13) ^b	2164	0.41	-0.10 (-0.11 to -0.09) ^b
		Participants	540 050 700	0.49	-0.02 (-0.06 to -0.01)	2 229 071	0.39	-0.12 (-0.15 to -0.08) ^b
Diabetes	0.48	Studies	3727	0.45	-0.03 (-0.03 to -0.02) ^b	1420	0.46	-0.03 (-0.03 to -0.02) ^b
		Participants	38 420 434	0.48	0.00 (-0.05 to 0.04)	4 823 058	0.47	-0.01 (-0.08 to 0.02)
Digestive	0.60	Studies	1282	0.49	-0.11 (-0.12 to -0.10) ^b	348	0.54	-0.06 (-0.08 to -0.04) ^b
		Participants	8 519 928	0.51	-0.09 (-0.13 to -0.07) ^b	147 821	0.56	-0.03 (-0.06 to -0.01)
Hepatitis A, B, C, and E	0.44	Studies	1131	0.34	-0.09 (-0.10 to -0.09) ^b	632	0.37	-0.06 (-0.07 to -0.05) ^b
		Participants	1 833 724	0.37	-0.06 (-0.17 to 0.06)	243 846	0.39	-0.05 (-0.07 to -0.03) ^b
HIV/AIDS	0.50	Studies	1741	0.33	-0.17 (-0.18 to -0.16) ^b	387	0.27	-0.23 (-0.25 to -0.21) ^b
		Participants	30 459 386	0.53	0.02 (-0.09 to 0.06)	155 531	0.35	-0.15 (-0.20 to -0.11) ^b
Kidney, chronic	0.57	Studies	2554	0.40	-0.17 (-0.17 to -0.16) ^b	476	0.42	-0.15 (-0.16 to -0.13) ^b
		Participants	18 747 970	0.44	-0.13 (-0.18 to -0.09) ^b	201 763	0.42	-0.15 (-0.17 to -0.12) ^b
Mental	0.48	Studies	3635	0.47	-0.01 (-0.02 to 0.00) ^b	1650	0.44	-0.04 (-0.05 to -0.03) ^b
		Participants	58 097 584	0.48	-0.01 (-0.19 to 0.07)	463 645	0.49	0.00 (-0.01 to 0.02)
Musculoskeletal	0.56	Studies	2418	0.66	0.10 (0.09 to 0.11) ^b	983	0.70	0.14 (0.13 to 0.15) ^b
		Participants	5 898 338	0.60	0.03 (0.00 to 0.08)	438 112	0.65	0.09 (-0.05 to 0.18)
Neoplasms	0.51	Studies	11 121	0.40	-0.11 (-0.11 to -0.11) ^b	3179	0.41	-0.10 (-0.11 to -0.10) ^b
		Participants	54 377 430	0.49	-0.03 (-0.04 to -0.01) ^b	2 946 236	0.50	-0.02 (-0.09 to 0.03)
Neurological	0.59	Studies	3431	0.50	-0.09 (-0.10 to -0.09) ^b	1338	0.52	-0.07 (-0.08 to -0.06) ^b
		Participants	10 576 242	0.53	-0.06 (-0.09 to -0.03) ^b	497 964	0.65	0.06 (-0.01 to 0.12)
Respiratory, chronic	0.48	Studies	2800	0.43	-0.04 (-0.05 to -0.04) ^b	1161	0.44	-0.03 (-0.04 to -0.02) ^b
		Participants	116 410 829	0.48	0.00 (-0.05 to 0.02)	1 231 162	0.47	-0.01 (-0.04 to 0.01)
Total ^c	0.54	Studies	48 211	0.42	-0.12 (-0.12 to -0.11) ^b	13 738	0.45	-0.09 (-0.09 to -0.08) ^b
		Participants	883 392 565	0.49	-0.05 (-0.06 to -0.03) ^b	13 378 210	0.48	-0.06 (-0.09 to -0.03) ^b

Abbreviation: AACT, Aggregate Analysis of ClinicalTrials.gov.

^a Data as of December 31, 2018. Published articles from 1966 to 2018 in PubMed were obtained using a search engine (Semantic Scholar)^{30,32}; clinical trial records from 1999 to 2018 were obtained from the AACT database.³³ Global prevalence data were obtained from the Global Health Data Exchange.³⁵ Sex bias with studies as measurement unit was defined as female participant fraction with studies as units (mean ratio of female participants/[male participants + female participants] for each study) minus female prevalence fraction, and is shown in rows with number of studies; sex bias with participants as measurement unit was defined as female participant fraction with participants as units (ratio of total number of female participants in all studies/total number of participants in all studies combined) minus female prevalence fraction, and is shown in rows with number of participants. Sex bias range was -1 to 1 ,

with 0 indicating no bias; negative sex bias indicates that female participants were represented less than male participants. Sex bias (1000 bootstrap estimates) is reported as mean and 95% bootstrap confidence interval (bottom 2.5%, top 97.5%).

^b Difference between sex bias value vs 0 : $P \leq .001$.

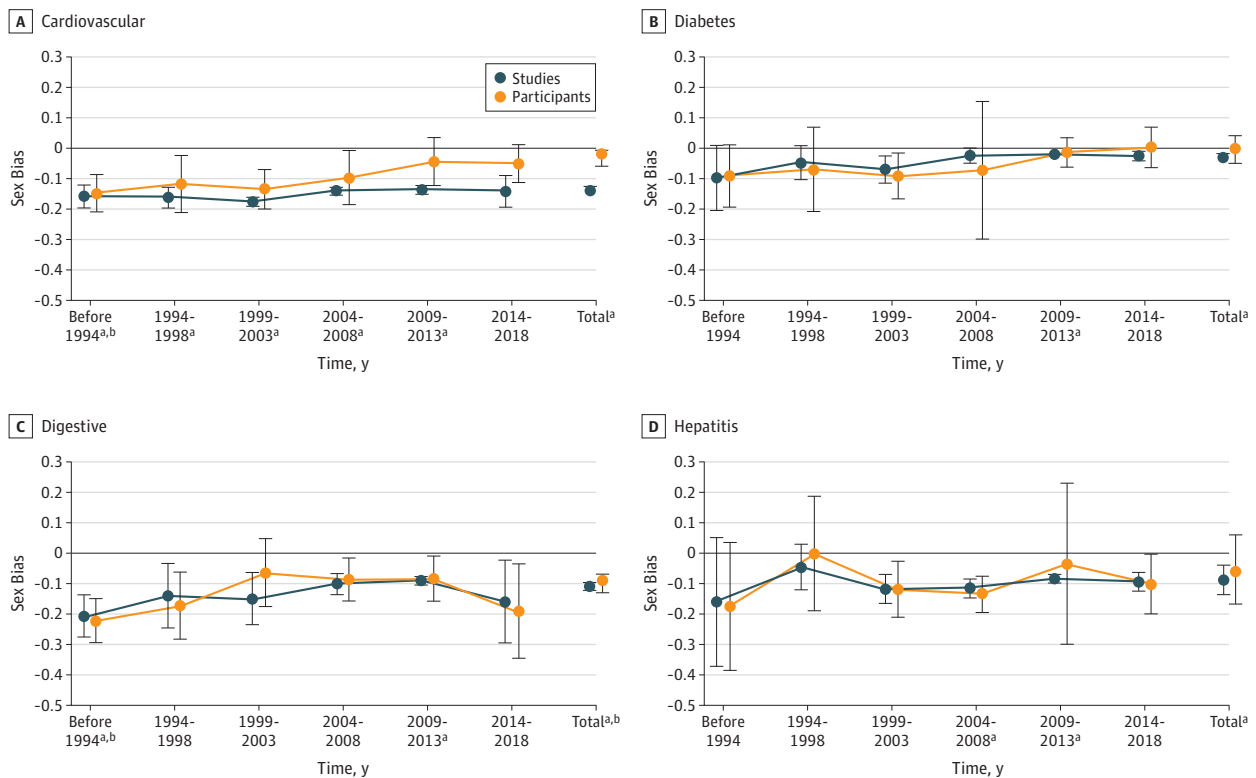
^c Totals include duplicate use of studies that mapped to more than 1 disease category. There were 38 506 of the 43 135 published articles (89%), representing 706 161 955 of the 792 004 915 participants (89%), and 12 609 of the 13 165 AACT records (96%), representing 12 636 768 of the 12 977 103 participants (97%), that mapped to a single disease category; only 4629 published articles (11%), representing 85 842 960 participants (11%), and 556 AACT records (4%), representing 340 335 participants (3%), contributed to sex bias estimates for more than 1 disease category.

about cardiovascular disease with participants as the measurement unit, mean sex bias was -0.02 [95% CI, -0.06 to -0.01]; with studies as the measurement unit, mean sex bias was -0.14 [95% CI, -0.14 to -0.13] (Table). Most articles and records mapped to a single disease category (Table).

With studies as measurement unit, sex bias was stable from before or during 1993 to 2018 for most disease categories (Figure 1, Figure 2, and Figure 3). With participants as measurement unit, sex bias improved (became less negative by ≥ 0.10) over time for cardiovascular diseases, HIV/AIDS, neoplasms, and neurological disorders (Figure 1, Figure 2, and Figure 3). Sex bias in articles for all categories combined was unchanged over time with studies as measurement unit (range, -0.15 [95% CI, -0.16 to -0.13] to -0.10 [95% CI, -0.14 to -0.06]), but improved from before 1993 (mean, -0.11 [95% CI, -0.16 to -0.05]) to 2014 to 2018 (mean -0.05 [95% CI, -0.09 to -0.02]) with participants as the measurement unit.

The mean absolute error between true F-Particip from data extracted manually vs automatically (PubMed-Extract) was 0.008. Errors made by PubMed-Extract were caused when (1) the table varied from typical table organization, (2) there were 2 or more columns for total counts and no single column for grand total, and (3) there were optical character recognition errors such as incorrect merging of multiple columns or splitting of single columns (eTable 4 in the Supplement). Manual analysis of automatically extracted participant numbers showed that 14 of 100 articles evaluated did not report the number of male and female participants, PubMed-Extract returned

Figure 1. Sex Bias in Clinical Studies Over Time Determined From Published Articles for Cardiovascular Diseases, Diabetes, Digestive Diseases, and Hepatitis (Types A, B, C, and E)



An intercept-only linear model was fitted to sex bias values from before and during 1993 and subsequently in 5-year increments. Estimated sex bias intercept coefficients were plotted against time for studies (blue) and participants as measurement unit (orange), with error bars representing 95% confidence intervals for the mean coefficients. The points for total at the right of each graph represent the mean sex bias totals for each category. Sex bias was defined as female participant fraction (determined separately for

studies and participants as measurement unit) minus female prevalence fraction (values for sex bias ranged from -1 to 1 , with 0 indicating no bias; negative sex bias indicates that female participants were represented less than male participants).

^a Difference between sex bias value vs 0 ; $P < .001$ for studies as measurement unit.

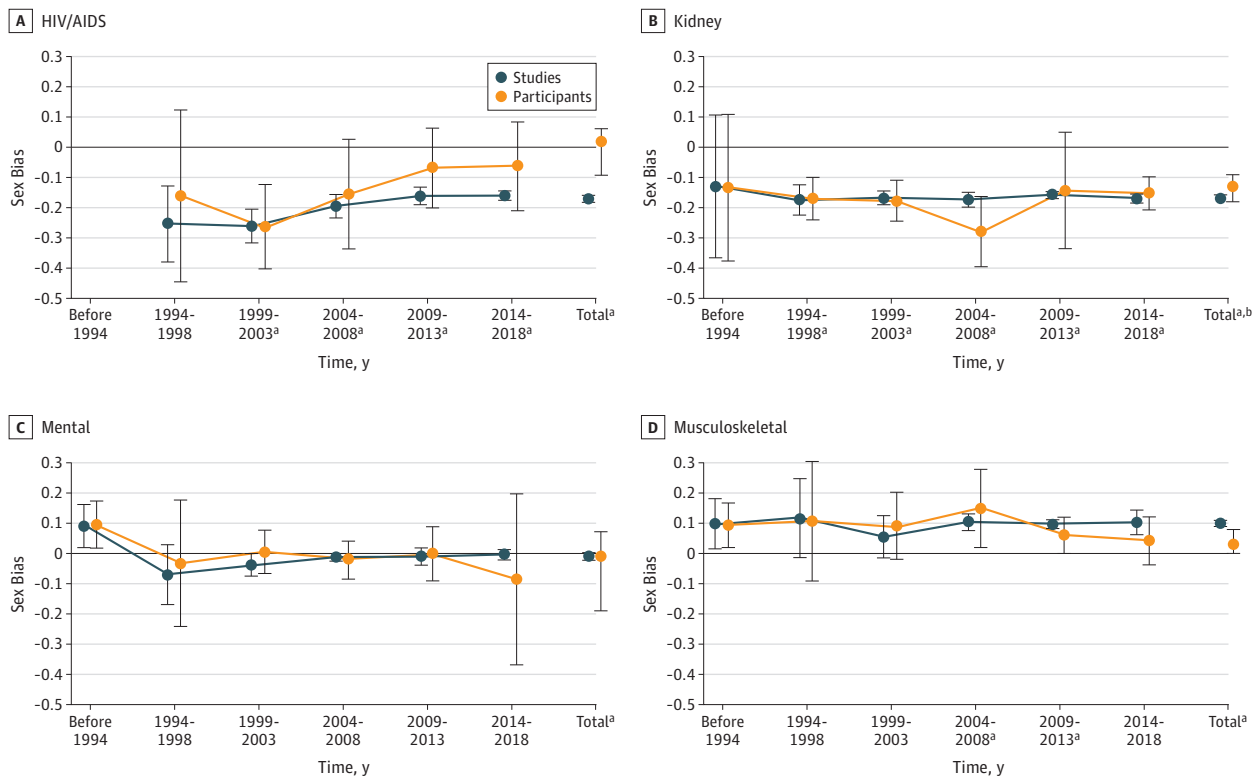
^b Difference between sex bias value vs 0 ; $P < .001$ for participants as measurement unit.

correct numerical estimates for 43 of the other 86 articles (recall, 50%), and mean precision for exact row extraction of male and female numbers was 0.75.

Comparison of the 1400 studies that had both articles and records showed that 675 studies (48%) had numbers of male and female participants that differed between articles and records, with magnitude of the difference between studies ranging from a minimum of 35 participants (52% of participants in the AACT record) to a maximum of 15 746 participants (92%). In 50 studies selected randomly from the 675 discordant studies, manual evaluation showed that discrepancies between articles and records were caused because the article was based on a subset of the trial data in the record (19 studies), PubMed-Extract extractions were incorrect or from the wrong table (14 studies), the article reported the number of participants who completed the trial vs the record that included enrolled participants who did not complete the trial (7 studies), the article was published before completion of the trial (3 studies), there was author error (1 study), and the article included patients from multiple trials (1 study); in 5 studies, the causes of discrepancies were unknown despite contacting authors for comments. In 6 of the 50 studies, the reasons for discrepancies were provided through email communication with study authors.

Linear regression with fixed effects to evaluate the association between publication year, disease category, and study size and sex bias in articles showed that the coefficients for number-of-participants deciles were positive and different from zero for the fifth decile (121-188 participants)

Figure 2. Sex Bias in Clinical Studies Over Time Determined From Published Articles for HIV/AIDS, Kidney Diseases (Chronic), Mental Disorders, and Musculoskeletal Disorders



An intercept-only linear model was fitted to sex bias values from before and during 1993 and subsequently in 5-year increments. Estimated sex bias intercept coefficients were plotted against time for studies (blue) and participants as measurement unit (orange), with error bars representing 95% confidence intervals for the mean coefficients. For HIV/AIDS before or during 1993, sex bias values for studies (-0.40) and participants (-0.42) were not plotted because they were based on only 3 articles (total, 138 participants). Sex bias was defined as female participant fraction (determined separately

for studies and participants as measurement unit) minus female prevalence fraction (values for sex bias ranged from -1 to 1, with 0 indicating no bias; negative sex bias indicates that female participants were represented less than male participants).

^a Difference between sex bias value vs 0; $P < .001$ for studies as measurement unit.

^b Difference between sex bias value vs 0; $P < .001$ for participants as measurement unit.

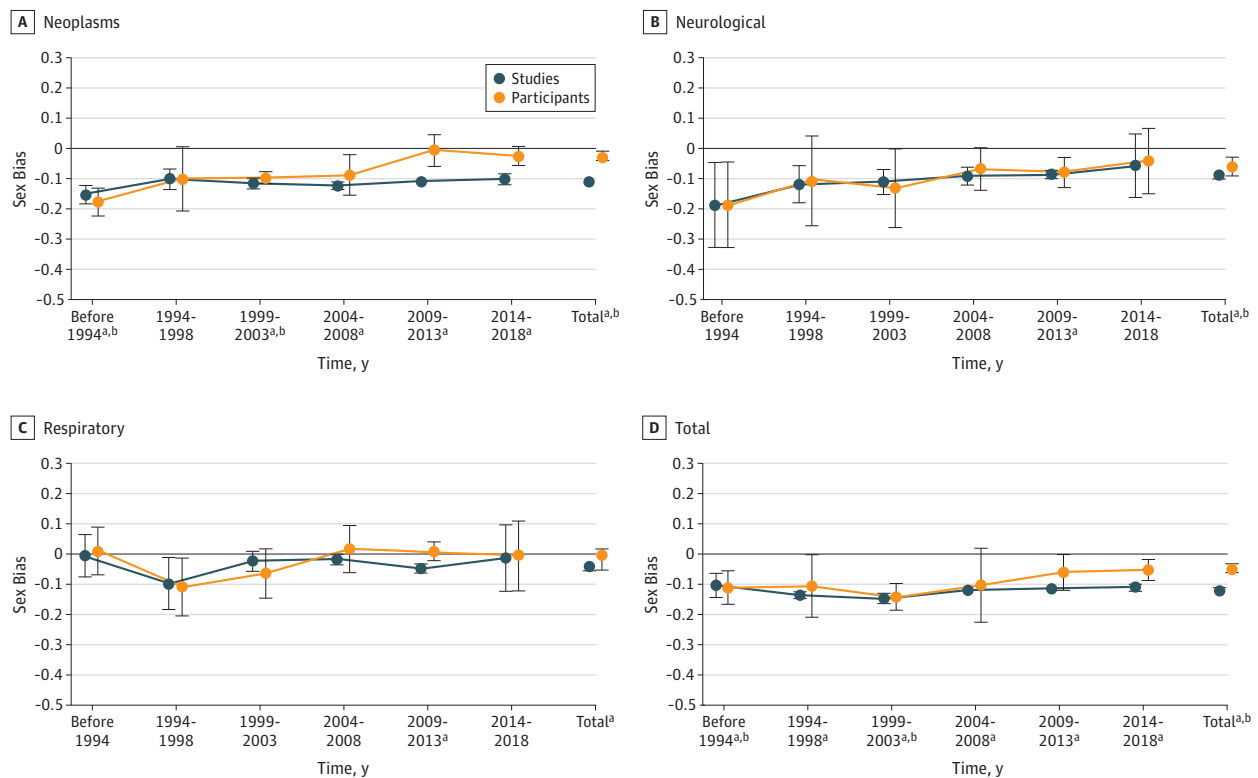
through 10th decile (≥ 2990 participants), indicating that larger study size was associated with greater female representation (eTable 6 in the Supplement).

Discussion

Using a large amount of data from articles and records, we observed substantial female underrepresentation in studies for diverse disease categories, especially HIV/AIDS and chronic kidney diseases. There was little increase in female representation in studies from before or during 1993 to 2018 using studies as measurement unit but improved female representation with participants as measurement unit (Figure 1, Figure 2, and Figure 3). Most disease categories were not evaluated previously (eTable 1 in the Supplement). The algorithms provided an effective and accurate automated scalable method for extracting male and female participant numbers and enabled expansion of analyses about sex bias to varied disease categories and integration of new data.

Previous studies of sex bias used studies or participants, but not both, as measurement unit (eTable 1 in the Supplement). With studies as measurement unit, each study has an equal contribution to the overall sex bias estimate, regardless of study size, providing a study-by-study evaluation of sex bias (Table, Figure 4). In contrast, with participants as measurement unit, participants may have an equal contribution to the overall sex bias estimate, providing a population

Figure 3. Sex Bias in Clinical Studies Over Time Determined From Published Articles for Neoplasms, Neurological Disorders, Respiratory Diseases (Chronic), and Total (All Categories Combined)



An intercept-only linear model was fitted to sex bias values from before and during 1993 and subsequently in 5-year increments. Estimated sex bias intercept coefficients were plotted against time for studies (blue) and participants as measurement unit (orange), with error bars representing 95% confidence intervals for the mean coefficients. The total number of published articles (all categories combined) increased from before or during 1993 (total, 482 articles) to 2014 to 2018 (18 627 articles). Sex bias in articles for all categories combined was unchanged over time with studies as measurement unit (range, $-0.15 [-0.16 \text{ to } -0.13]$ to $-0.10 [-0.14 \text{ to } -0.06]$), but improved from before 1993

($-0.11 [-0.16 \text{ to } -0.05]$) to 2014 to 2018 ($-0.05 [-0.09 \text{ to } -0.02]$) with participants as measurement unit. Sex bias was defined as female participant fraction (determined separately for studies and participants as measurement unit) minus female prevalence fraction (values for sex bias ranged from -1 to 1 , with 0 indicating no bias; negative sex bias indicates that female participants were represented less than male participants).

^a Difference between sex bias value vs 0 ; $P < .001$ for studies as measurement unit.

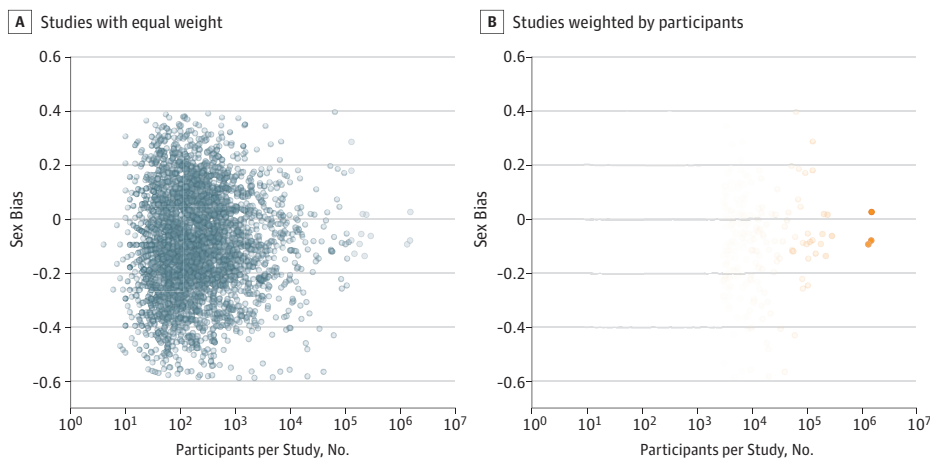
^b Difference between sex bias value vs 0 ; $P < .001$ for participants as measurement unit.

estimate; however, larger studies contribute proportionally more, and smaller studies have a nearly invisible contribution to overall sex bias estimates (Figure 4). The marked difference in sex bias in articles with studies vs participants as measurement unit for cardiovascular diseases (−0.14 vs −0.02) and neoplasms (−0.11 vs −0.03) is evidence that sex bias determined with both measurement units should be reported, and that sex bias results may be less sensitive to female underrepresentation with participants than studies as measurement unit (Table, Figure 1, Figure 2, and Figure 3). The use of studies as measurement unit may ensure that small studies of less prevalent diseases receive equal representation in estimates of overall sex bias (Figure 4). The limited change in sex bias over time for all categories combined with studies as measurement unit (Figure 3) may be addressed with policy and funding initiatives that focus on sex bias regardless of proposed study size. Furthermore, the importance of study size was underscored by the relation between study size and female representation in articles (eTable 6 in the Supplement).

With studies as measurement unit, sex bias estimates from articles and records were consistent in polarity and magnitude for diabetes, HIV/AIDS, kidney diseases, mental disorders, neoplasms, neurological disorders, and respiratory diseases but differed in magnitude for digestive diseases and musculoskeletal disorders (Table). Differences in sex bias estimates may, in part, be due to having fewer records than articles (digestive diseases, 348 records vs 1282 articles), and AACT data may have been biased geographically because trial registration requirements for ClinicalTrials.gov may apply only to US clinical trials.³⁴ Geographic differences may be important because of marked variation in regional disease profiles, such as differences in HIV/AIDS incidence between sub-Saharan Africa vs East Asia.³⁸ Future studies may include machine reading algorithms to evaluate study location.

Differences in sex bias estimates between articles vs records also may be due to discrepancies in male and female participant numbers between articles and records observed in 48% of studies. Manual evaluation of these discrepancies was limited to 50 studies because it was time-consuming and associated with delays inherent with email queries to authors when reasons for discrepancies could not be ascertained from the article and record. A previous comparison of randomized clinical drug trials in ClinicalTrials.gov vs counterpart published articles concluded that trial results should be evaluated systematically from both sources because of important differences, including more complete reporting in records than articles, variation in reporting between articles from specialty vs general journals, and absence of an article corresponding to 50% of trials posted on ClinicalTrials.gov (so-called abandoned trials).^{39,40} Trial registration and reporting on ClinicalTrials.gov may vary between studies funded by industry or government sources, and the requirement of mandatory

Figure 4. Sex Bias vs Number of Study Participants for 14 371 Cardiovascular Clinical Studies, Estimated From Published Articles by the PubMed-Extract Algorithm



Each point represents 1 article. A, With studies as the measurement unit of sex bias, each study point has equal intensity of blue shade and contribution to the overall estimate of sex bias. B, With participants as the measurement unit of sex bias, study point orange shade intensity is proportional to the number of participants; small studies are essentially invisible and contribute little to the overall sex bias estimate.

posting of trial results on ClinicalTrials.gov within 1 year of completion of data collection is adhered to infrequently and may promote the posting of cursory reports that may include inaccurate or incomplete data that are not peer reviewed.^{6,41-43} Journal publication may be associated with partial and altered reporting (so-called filtered data) due to space limitations, publication bias, revised analyses and data exclusion due to suggestions from peer reviewers, and delays inherent in journal submission and peer review.^{40,41} The observation of sex bias differences between articles and records is further evidence to support the need for greater transparency and accuracy in trial reporting in both media.

The comparison of data from articles vs records may have been affected by our decision to include data from articles about studies other than clinical trials, such as observational studies, case series studies, and quality improvement analyses. Although a focus on trials alone may provide a more direct comparison between data from articles vs records, the inclusion of all published articles may provide a more realistic description of current sex bias in funded and nonfunded clinical research. Observational studies may be considered lower in evidence quality than trials but remain important because they provide valuable context for trial results and data in areas with limited trials.⁴⁴⁻⁴⁶ Furthermore, randomized trials may not necessarily represent general disease populations because of participant exclusion criteria.⁴⁷ Nevertheless, sex bias estimates for trials alone may be determined in future work by applying different filters to the data extraction algorithms.

In selecting disease categories that previously were defined in GHDx, we recognized potential overlap between categories, such as cardiovascular, kidney, or neurological diseases in studies of patients who had diabetes. Nevertheless, the disease categories were used because they represented large, important, clinically relevant categories. Most studies were limited to only 1 of the 11 disease categories, and only 11% of articles and 4% of records contributed to sex bias estimates for more than 1 disease category (Table). The attribution of cost and resource allocation to overlapping disease categories is an inherent issue in epidemiology and public health that we addressed by specifying the sources of disease category definitions and data and quantifying the number of studies that mapped to more than 1 category.⁴⁸

Limitations

Limitations of the present study include the analysis of sex bias without other variables. Sex bias may vary with age for colorectal and lung cancer¹²; further evaluation using our algorithms may enable robust analysis of the interaction between sex, age, and race in study enrollment. We did not evaluate diagnoses that have marked variation of sex prevalence within disease categories, such as different types of cancer (eg, breast vs prostate cancer), because our goal was to provide a broad overview about sex bias for different disease categories; in future work, filters added to the data extraction code may enable more focused sex bias data for specific diseases. In addition, we included participant counts from primary studies and secondary analyses such as meta-analyses and systematic reviews, but in estimating sex bias, we did not account for multiple inclusion of the same primary study participants in the secondary analyses; therefore, estimates of sex bias from articles may have been affected preferentially by primary studies that were included in secondary analyses, and the magnitude of this effect is unknown. The total number of more than 792 million participants may seem unrealistically high because it may imply that 10% of the 7.7 billion people globally were involved in a clinical study; the large number of participants may have been affected by large population-based studies including a survey from China (381 million participants) and study of death records from the United States, England, and Wales (almost 86 million participants) that accounted for 467 million participants (53%).^{49,50} In future big data studies that are based on articles, it may be advisable to modify the data extraction coding to exclude duplicate use of studies and analyze large outlier studies separately. For the time series, we used publication date of articles and did not extract information about the time range of study execution; that may be considered in future work.

Conclusions

Automated extraction of participant numbers in clinical reports provides an effective alternative to manual analysis of demographic bias and may expedite analyses for multiple diseases globally. Our findings indicate that studies with more participants have greater female representation. However, sex bias against female participants in clinical studies persists despite legal and policy initiatives to increase female representation.

ARTICLE INFORMATION

Accepted for Publication: May 17, 2019.

Published: July 3, 2019. doi:10.1001/jamanetworkopen.2019.6700

Open Access: This is an open access article distributed under the terms of the [CC-BY License](#). © 2019 Feldman S et al. *JAMA Network Open*.

Corresponding Author: Sergey Feldman, PhD, Allen Institute for Artificial Intelligence, 2157 N Northlake Way, Ste 110, Seattle, WA 98103 (sergey@allenai.org).

Author Affiliations: Allen Institute for Artificial Intelligence, Seattle, Washington (Feldman, Ammar, Lo, Trepman, van Zuylen, Etzioni); University of South Alabama College of Medicine, Mobile (Trepman); Department of Medical Microbiology and Infectious Diseases, University of Manitoba, Winnipeg, Manitoba, Canada (Trepman).

Author Contributions: Dr Feldman had full access to all of the data in the study and takes responsibility for the integrity of the data and the accuracy of the data analysis.

Concept and design: Feldman, Ammar, Etzioni.

Acquisition, analysis, or interpretation of data: Feldman, Ammar, Lo, Trepman, van Zuylen.

Drafting of the manuscript: Feldman, Ammar, Lo, Trepman, van Zuylen.

Critical revision of the manuscript for important intellectual content: Feldman, Ammar, Lo, Trepman, Etzioni.

Statistical analysis: Feldman, Lo.

Administrative, technical, or material support: Ammar, Trepman, Etzioni.

Supervision: Feldman, Ammar, Etzioni.

Conflict of Interest Disclosures: Dr Feldman reported serving as a consultant for the Bill & Melinda Gates Foundation outside the submitted work. No other disclosures were reported.

Funding/Support: This study was funded by the Allen Institute for Artificial Intelligence.

Role of the Funder/Sponsor: The authors did this work as part of their work duties at the Allen Institute for Artificial Intelligence, including the design and conduct of the study; collection, management, analysis, and interpretation of the data; preparation, review, or approval of the manuscript; and decision to submit the manuscript for publication.

Additional Contributions: We thank the Clinical Trials Transformation Initiative for access to the Aggregate Analysis of ClinicalTrials.gov database and the Institute for Health Metrics and Evaluation for the Global Health Data Exchange. We also thank the authors of studies who were contacted to help determine reasons for discrepancies between articles and records. Ruth Etzioni, PhD (Fred Hutchinson Cancer Research Center), Brad H. Pollock, MPH, PhD (University of California Davis), Margaret Rosenfeld, MD, MPH (University of Washington), Lucy Lu Wang, MS (University of Washington), and Dan Weld, PhD (University of Washington and Allen Institute), performed manuscript review and provided comments. Luca Weihs, BA (University of Washington), contributed to helpful discussion about statistical methods. David Orentlicher, MD, JD (UNLV William S. Boyd School of Law), Craig Shapiro, MD (retired, US Public Health Service), and Carole Stipelman, MD, MPH (University of Utah School of Medicine), provided helpful discussion and comments. No compensation was received by the acknowledged individuals.

REFERENCES

1. Wallach JD, Sullivan PG, Trepanowski JF, Steyerberg EW, Ioannidis JP. Sex based subgroup differences in randomized controlled trials: empirical evidence from Cochrane meta-analyses. *BMJ*. 2016;355:i5826. doi:10.1136/bmj.i5826
2. Whitley H, Lindsey W. Sex-based differences in drug activity. *Am Fam Physician*. 2009;80(11):1254-1258.

3. Heinrich J. Drug safety: most drugs withdrawn in recent years had greater health risks for women. <https://www.gao.gov/assets/100/90642.pdf>. Published January 19, 2001. Accessed November 10, 2018.
4. McGregor AJ. Sex bias in drug research: a call for change. *Pharm J*. 2016;296(7887). <https://www.pharmaceutical-journal.com/opinion/comment/sex-bias-in-drug-research-a-call-for-change/20200727.article>. Published March 16, 2016. Accessed November 9, 2018.
5. Farkas RH, Unger EF, Temple R. Zolpidem and driving impairment—identifying persons at risk. *N Engl J Med*. 2013;369(8):689-691. doi:10.1056/NEJMp1307972
6. Food and Drug Administration Amendments Act of 2007, Pub L No. 110-85, 121 stat 823, 110th Cong. <https://www.gpo.gov/fdsys/pkg/PLAW-110publ85/pdf/PLAW-110publ85.pdf>. Accessed November 30, 2018.
7. Tran C, Knowles SR, Liu BA, Shear NH. Gender differences in adverse drug reactions. *J Clin Pharmacol*. 1998;38(11):1003-1009. doi:10.1177/009127009803801103
8. Zopf Y, Rabe C, Neubert A, et al. Women encounter ADRs more often than do men. *Eur J Clin Pharmacol*. 2008;64(10):999-1004. doi:10.1007/s00228-008-0494-6
9. Weisman CS, Cassard SD. Health consequences of exclusion or underrepresentation of women in clinical studies. In: Mastroianni AC, Faden R, Federman D, eds. *Women and Health Research: Ethical and Legal Issues of Including Women in Clinical Studies*. Vol 2. Washington, DC: National Academies Press; 1994:35-40.
10. National Institutes of Health Revitalization Act of 1993. Subtitle B—clinical research equity regarding women and minorities. <https://orwh.od.nih.gov/sites/orwh/files/docs/NIH-Revitalization-Act-1993.pdf>. Accessed November 9, 2018.
11. Ramasubbu K, Gurm H, Litaker D. Gender bias in clinical trials: do double standards still apply? *J Womens Health Gen Based Med*. 2001;10(8):757-764. doi:10.1089/15246090152636514
12. Murthy VH, Krumholz HM, Gross CP. Participation in cancer clinical trials: race-, sex-, and age-based disparities. *JAMA*. 2004;291(22):2720-2726. doi:10.1001/jama.291.22.2720
13. Hutchins LF, Unger JM, Crowley JJ, Coltman CA Jr, Albain KS. Underrepresentation of patients 65 years of age or older in cancer-treatment trials. *N Engl J Med*. 1999;341(27):2061-2067. doi:10.1056/NEJM199912303412706
14. Geller SE, Adams MG, Carnes M. Adherence to federal guidelines for reporting of sex and race/ethnicity in clinical trials. *J Womens Health (Larchmt)*. 2006;15(10):1123-1131. doi:10.1089/jwh.2006.15.1123
15. Geller SE, Koch A, Pellettieri B, Carnes M. Inclusion, analysis, and reporting of sex and race/ethnicity in clinical trials: have we made progress? *J Womens Health (Larchmt)*. 2011;20(3):315-320. doi:10.1089/jwh.2010.2469
16. Harris DJ, Douglas PS. Enrollment of women in cardiovascular clinical trials funded by the National Heart, Lung, and Blood Institute. *N Engl J Med*. 2000;343(7):475-480. doi:10.1056/NEJM200008173430706
17. Hoel AW, Kayssi A, Brahmanandam S, Belkin M, Conte MS, Nguyen LL. Under-representation of women and ethnic minorities in vascular surgery randomized controlled trials. *J Vasc Surg*. 2009;50(2):349-354. doi:10.1016/j.jvs.2009.01.012
18. Ibrahim M, Ogunleye F, Roye J, Yadav S, Townsel D, Yu Z. Representation of minorities and elderly in cancer clinical trials at a single institution—the William Beaumont Hospital experience. *J Cancer Epidemiol Prev*. 2017;2(1):1.
19. Kalliainen LK, Wisecarver I, Cummings A, Stone J. Sex bias in hand surgery research. *J Hand Surg Am*. 2018;43(11):1026-1029. doi:10.1016/j.jhsa.2018.03.026
20. Klabunde CN, Springer BC, Butler B, White MS, Atkins J. Factors influencing enrollment in clinical trials for cancer treatment. *South Med J*. 1999;92(12):1189-1193. doi:10.1097/00007611-199912000-00011
21. Polit DF, Beck CT. Is there still gender bias in nursing research? an update. *Res Nurs Health*. 2013;36(1):75-83. doi:10.1002/nur.21514
22. Robbins NM, Bernat JL. Minority representation in migraine treatment trials. *Headache*. 2017;57(3):525-533. doi:10.1111/head.13018
23. Stewart JH, Bertoni AG, Staten JL, Levine EA, Gross CP. Participation in surgical oncology clinical trials: gender-, race/ethnicity-, and age-based disparities. *Ann Surg Oncol*. 2007;14(12):3328-3334. doi:10.1245/s10434-007-9500-y
24. Vidaver RM, Lafleur B, Tong C, Bradshaw R, Marts SA. Women subjects in NIH-funded clinical research literature: lack of progress in both representation and analysis by sex. *J Womens Health Gen Based Med*. 2000;9(5):495-504. doi:10.1089/15246090050073576
25. Ashish N, Patawari A. Machine reading of biomedical data dictionaries. *ACM J Data Inf Qual*. 2018;9(4):21. doi:10.1145/3177874

26. Tsutsui S, Ding Y, Meng G. Machine reading approach to understand Alzheimer's disease literature. Paper presented at: Conference on Information and Knowledge Management; Indianapolis, IN; October 24-28, 2016. http://homes.sice.indiana.edu/stsutsui/pub_pdfs/machine_reading_ad.pdf. Accessed December 9, 2018.
27. Šuster S, Daelemans W. ClICR: a dataset of clinical case reports for machine reading comprehension. Paper presented at: North American Chapter of the Association for Computational Linguistics: Human Language Technologies; New Orleans, LA; June 1-6, 2018. <https://arxiv.org/pdf/1803.09720.pdf>. Accessed December 9, 2018.
28. Cohen PR. DARPA's Big Mechanism program. *Phys Biol*. 2015;12(4):045008. doi:10.1088/1478-3975/12/4/045008
29. Etzioni O, Banko M, Cafarella MJ. Machine reading. In: Cohn A, ed. *Proceedings of the 21st National Conference on Artificial Intelligence, Boston, Massachusetts—July 16-20, 2006*. Vol 2. Palo Alto, CA: AAAI Press; 2006:1517-1519. <https://www.aaai.org/Papers/AAAI/2006/AAAI06-239.pdf>. Accessed December 9, 2018.
30. Allen Institute for Artificial Intelligence. Semantic Scholar. <https://allenai.org/semantic-scholar/>. Accessed November 11, 2018.
31. Bhagavatula C, Feldman S, Power R, Ammar W. Content-based citation recommendation. Paper presented at: 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies; New Orleans, LA; June 1-6, 2018. <http://aclweb.org/anthology/N18-1022>. Accessed November 9, 2018.
32. US National Library of Medicine. PubMed. <https://www.ncbi.nlm.nih.gov/pubmed/>. Accessed November 11, 2018.
33. Aggregate Analysis of ClinicalTrials.gov database. <https://www.ctti-clinicaltrials.org/aact-database>. Accessed November 11, 2018.
34. US National Library of Medicine. FDAAA 801 and the Final Rule. <https://clinicaltrials.gov/ct2/manage-recs/fdaaa>. Accessed November 29, 2018.
35. Institute for Health Metrics and Evaluation. Global Health Data Exchange. <http://ghdx.healthdata.org/gbd-results-tool>. Accessed November 11, 2018.
36. GBD 2016 Causes of Death Collaborators. Global, regional, and national age-sex specific mortality for 264 causes of death, 1980-2016: a systematic analysis for the Global Burden of Disease Study 2016. *Lancet*. 2017;390(10100):1151-1210. doi:10.1016/S0140-6736(17)32152-9
37. US National Library of Medicine. MEDLINE®PubMed® XML element descriptions and their attributes: 24: <PublicationTypeList>. https://www.nlm.nih.gov/bsd/licensee/elements_descriptions.html#publicationtypelist. Accessed November 29, 2018.
38. Fattig J, Swaminathan M, Murrill CS, Kaplan JE. Global epidemiology of HIV. *Infect Dis Clin North Am*. 2014;28(3):323-337. doi:10.1016/j.idc.2014.05.001
39. Riveros C, Dechartres A, Perrodeau E, Haneef R, Boutron I, Ravaud P. Timing and completeness of trial results posted at ClinicalTrials.gov and published in journals. *PLoS Med*. 2013;10(12):e1001566. doi:10.1371/journal.pmed.1001566
40. Doshi P, Dickersin K, Healy D, Vedula SS, Jefferson T. Restoring invisible and abandoned trials: a call for people to publish the findings. *BMJ*. 2013;346:f2865. doi:10.1136/bmj.f2865
41. Choi R. Increasing transparency of clinical trial data in the United States and the European Union. *Wash Univ Glob Stud Law Rev*. 2015;14(3):521-548.
42. Law MR, Kawasumi Y, Morgan SG. Despite law, fewer than one in eight completed studies of drugs and biologics are reported on time on ClinicalTrials.gov. *Health Aff (Millwood)*. 2011;30(12):2338-2345. doi:10.1377/hlthaff.2011.0172
43. Zarin DA, Tse T, Williams RJ, Rajakannan T. Update on trial registration 11 years after the ICMJE policy was established. *N Engl J Med*. 2017;376(4):383-391. doi:10.1056/NEJMs1601330
44. Barnish MS, Turner S. The value of pragmatic and observational studies in health care and public health. *Pragmat Obs Res*. 2017;8:49-55. doi:10.2147/POR.S137701
45. Cole AP, Abdollah F, Trinh QD. Observational studies to contextualize surgical trials. *Eur Urol*. 2016;70(2):231-232. doi:10.1016/j.eururo.2016.02.062
46. Dreyer NA, Tunis SR, Berger M, Ollendorff D, Mattox P, Glick R. Why observational studies should be among the tools used in comparative effectiveness research. *Health Aff (Millwood)*. 2010;29(10):1818-1825. doi:10.1377/hlthaff.2010.0666
47. Kennedy-Martin T, Curtis S, Faries D, Robinson S, Johnston J. A literature review on the representativeness of randomized controlled trial samples and implications for the external validity of trial results. *Trials*. 2015;16:495. doi:10.1186/s13063-015-1023-4

48. Nichols GA, Brown JB. The impact of cardiovascular disease on medical care costs in subjects with and without type 2 diabetes. *Diabetes Care*. 2002;25(3):482-486. doi:10.2337/diacare.25.3.482
49. Wang Z, Cao C, Guo C, Chen G, Chen H, Zheng X. Socioeconomic inequities and cardiovascular disease-related disability in China: a population-based study. *Medicine (Baltimore)*. 2016;95(32):e4409. doi:10.1097/MD.0000000000004409
50. Hurley MN, McKeever TM, Prayle AP, Fogarty AW, Smyth AR. Rate of improvement of CF life expectancy exceeds that of general population—observational death registration study. *J Cyst Fibros*. 2014;13(4):410-415. doi:10.1016/j.jcf.2013.12.002

SUPPLEMENT.

eAppendix. Quantifying Sex Bias in Clinical Studies at Scale With Automated Data Extraction

eReferences

eTable 1. Previous Studies About Enrollment Sex Bias in Clinical Research

eTable 2. Global Disease Category Prevalence

eTable 3. Categories in the XML PubMed Field <PublicationTypeList> Used to Identify Study-Related Articles in PubMed

eTable 4. Example Table Used for Illustration of Sex Data Extraction Using PubMed-Extract

eTable 5. Relation Between Disease Category and Medical Subject Headings Terms

eTable 6. Linear Model for Number of Participants vs Enrollment Sex Bias, Controlling for Publication Year and Disease Category