

1 **Inferring Gender from First Names: Comparing the Accuracy of Genderize, Gender API, and the**
2 **gender R Package on Authors of Diverse Nationality**

3
4 Alexander D. VanHelene, BS^{1,2}; Ishaani Khatri, BS³, C. Beau Hilton, MD⁴, Sanjay Mishra, MS, PhD^{1,2,3},
5 Ece D. Gamsiz Uzun, MS, PhD, FAMIA^{2,3,5,6}, Jeremy L. Warner, MD, MS, FAMIA, FASCO^{1,2,3*}

6
7 ¹Lifespan Cancer Institute, Rhode Island Hospital, Providence, Rhode Island

8 ²Center for Clinical Cancer Informatics and Data Science, Legorreta Cancer Center, Brown University,
9 Providence, Rhode Island

10 ³Warren Alpert Medical School, Brown University, Providence, Rhode Island

11 ⁴Department of Internal Medicine, Vanderbilt University, Nashville, Tennessee

12 ⁵Center for Computational Molecular Biology, Brown University, Providence, Rhode Island

13 ⁶Department of Pathology and Laboratory Medicine, Brown University, Providence, Rhode Island

14
15 *Corresponding Author:

16 Jeremy L. Warner, MD, MS, FAMIA, FASCO

17 593 Eddy St.

18 Providence, RI 02903

19 401-444-7882

20 jeremy_warner@brown.edu

21

22

23

24

25

26

27 **Abstract**

28 Meta-researchers commonly leverage tools that infer gender from first names, especially when
29 studying gender disparities. However, tools vary in their accuracy, ease of use, and cost. The objective of
30 this study was to compare the accuracy and cost of the commercial software Genderize and Gender API,
31 and the open-source gender R package. Differences in binary gender prediction accuracy between the
32 three services were evaluated. Gender prediction accuracy was tested on a multi-national dataset of
33 32,968 gender-labeled clinical trial authors. Additionally, two datasets from previous studies with 5779
34 and 6131 names, respectively, were re-evaluated with modern implementations of Genderize and Gender
35 API. The gender inference accuracy of Genderize and Gender API were compared, both with and without
36 supplying trialists' country of origin in the API call. The accuracy of the gender R package was only
37 evaluated without supplying countries of origin since. The accuracy of Genderize, Gender API, and the
38 gender R package were defined as the percentage of correct gender predictions. Accuracy differences
39 between methods were evaluated using McNemar's test. Genderize and Gender API demonstrated overall
40 96.6% and 96.1% accuracy, respectively, when countries of origin were not supplied in the API calls.
41 Genderize and Gender API achieved the highest accuracy when predicting the gender of German authors
42 with accuracies greater than 98%. Genderize and Gender API were least accurate with South Korean,
43 Chinese, Singaporean, and Taiwanese authors, demonstrating below 82% accuracy. The gender R
44 package achieved below 86% accuracy on the full dataset. In the replication studies, Genderize and
45 gender API demonstrated better performance than in the original publications. Our results indicate that
46 Genderize and Gender API are highly accurate, except when evaluating South Korean, Chinese,
47 Singaporean, and Taiwanese names. We also demonstrated that Genderize can provide similar accuracy
48 to Gender API while being 4.85x less expensive.

49

50

51

52

53 **Author Summary**

54 Gender disparities in academia have prompted researchers to investigate gender gaps in
55 professorship roles and publication authorship. Of particular concern are the gender gaps in cancer
56 clinical trial authorship. Methodologies that evaluate gender disparities in academia often rely on tools
57 that infer gender from first names. Tools that predict gender from first names are often used in
58 methodologies that determine the gender ratios of academic departments or publishing authors in a
59 discipline. However, researchers must choose between different gender predicting tools that vary in their
60 accuracy, ease of use, and cost. We evaluated the binary gender prediction accuracy of Genderize, Gender
61 API, and the gender R package on a gold-standard dataset of 32,968 clinical trialists from around the
62 world. Genderize and Gender API cost money to use, while the gender R package is free and open source.
63 We found that Genderize and Gender API were more accurate than the gender R package. In addition,
64 Genderize is cheaper than Gender API, but is more sensitive to inconsistencies in name formatting and
65 the presence of diacritical marks. Both Genderize and Gender API were most accurate with western
66 names.

68 **Introduction**

69 One of the most well-documented disparities in STEM is gender disparity [1,2]. This issue is
70 especially notable in the cancer clinical trial domain, with underrepresentation of women in the leadership
71 of pivotal trials documented as recently as within the last decade [3]. The study of gender disparity in
72 scientific authorship often requires the determination of gender from very limited data, e.g., author
73 forenames. Software [4–9] that infers gender from forenames could potentially enable researchers to
74 automate gender prediction in large datasets. Commercial gender prediction services [10,11] such as
75 Genderize and Gender API programmatically predict gender from first names. The gender R package [12]
76 is an open-source alternative to these proprietary gender prediction tools.

77 Gender prediction software has demonstrated high accuracy when evaluating Western first
78 names, but often falters when evaluating names from Asian cultures [13]. Further, the presence of

79 diacritical marks and hyphens reportedly affects the accuracy of gender prediction in some tools [14].
80 Few studies [15] to date have evaluated differences in accuracy in gender predicting software between
81 Western and non-Western names. To our knowledge, no studies have evaluated how different ways of
82 delimiting two-part first names e.g. Jean-Pierre vs Jean Pierre vs Jeanpierre, affect gender prediction
83 accuracy.

84 We compared the gender prediction accuracy of Genderize, Gender API, and the gender R
85 package using a large manually curated registry of cancer clinical trialists with labeled genders and
86 diverse nationalities. In addition, we quantified the accuracy of these tools by author nationality and
87 compared different strategies for delimiting two-part forenames, which are common in the English
88 language spelling of Korean, Chinese, Singaporean, and Taiwanese names.

89

90 **Materials and Methods**

91 Three gender prediction tools: 1) Genderize; 2) Gender API; and 3) the gender R package, were
92 tested on a gold-standard registry of cancer clinical trialists with manually determined binary gender.
93 Trialists' names and affiliations were sourced from the HemOnc knowledge base, [16] a continually
94 growing resource created to capture the standard-of-care treatments in the fields of hematology and
95 oncology. The binary gender classifications used in our study refer to socially constructed gender
96 categories, not biological sex [17,18]. Names in HemOnc are primarily sourced from the MEDLINE
97 records of published clinical trials and undergo extensive normalization to account for the presence of
98 diacritics, middle initials, misspellings, multipart last names represented as middle names, and other
99 variations. When first names are not available through MEDLINE, the original manuscripts are examined
100 for this information. Binary gender is determined by a combination of automated mappings of typically
101 masculine or feminine forenames (e.g., John; Rebecca), web searches of publicly available information
102 such as biographies on academic web pages, and consensus determinations including consultation with
103 native speakers. If gender cannot be determined after these efforts, the author is labeled as “unknown

104 gender”. A subset of journals does not provide forenames; in these cases, the gender is labeled as “could
105 not be determined.” Country affiliations sourced from MEDLINE also undergo extensive normalization.

106 Gender prediction accuracy was defined as the percent of individuals whose gender was correctly
107 predicted, as compared to the gold standard dataset. The percent of incorrect gender predictions and the
108 percent of names with no predicted gender were also calculated. For binary statistical tests, gender
109 predictions were categorized as successes or failures – correct gender predictions were defined as
110 successes, while names with incorrect or absent predictions were failures.

111 All trialists with a gender determination were evaluated with Genderize and Gender API on 2023-
112 11-21 using the R package httr (version 1.4.7). Both US Social Security Administration (SSA) and US
113 Census Integrated Public Use Microdata Series (IPUMS) name datasets were used as a reference when
114 predicting names with the gender R package [12] (version 0.6.0).

115 Genderize and Gender API were used to predict names with and without supplying a country of
116 origin for the subset of authors with a singular country of affiliation. The Gender R package was only
117 tested without supplying country names because the SSA and IPUMS methods do not provide that
118 functionality. Two-part names were concatenated without any delimiter e.g. Jean-Pierre was converted to
119 jeanpierre. Middle names were removed, unless an author had a first initial/middle name, in which case
120 their middle name was used. Gender bias in name prediction was descriptively evaluated by calculating
121 the percent of names that were misgendered, compared to the gold standard labeled dataset. In an
122 additional analysis, accuracy differences resulting from delimiting two-part first names with different
123 characters were evaluated. Two-part first name prediction accuracy was also evaluated using the first half
124 of two-part names only. For example, the name Jean-Pierre was tested four ways: 1) jean-pierre; 2) jean
125 pierre; 3) jeanpierre; and 4) jean.

126 In addition to predicting the gender of a first name, Genderize and Gender API also report an
127 estimated probability that a gender prediction is correct. We evaluated the correlation between these API-
128 reported probability estimates and the gold standard labeled dataset with linear regressions and Brier

129 scores. Names with a reported probability less than or equal to 50% were excluded from the regression
130 and Brier scores.

131 The gender prediction accuracies of Genderize and Gender API were also separately evaluated
132 using publicly available datasets from two studies [15,19] that tested gender prediction in 2018 and 2021,
133 respectively. The dataset [20] provided by Santamaria 2018 consisted of 5,779 names sourced from
134 various other datasets. The dataset [21] sourced from Sebo 2021 consisted of 6,131 Swiss physicians. The
135 names from these public datasets were not modified prior to our evaluation on 2023-11-07. Nor were
136 nationalities supplied to Genderize and Gender API when evaluating these public datasets, following the
137 original experimental design.

138 All software accuracy comparisons were computed in R version 4.3.1. Differences in accuracy
139 between methods were evaluated using the default R stats package implementation of McNemar's test
140 [22]. Data analysis was facilitated with tidyverse [23] (version 2.0.0), haven [24] (version 2.5.3), readxl
141 [25] (version 1.4.3), testthat [26] (version 3.1.10), ggpmisc [27] (version 0.5.5), and patchwork [28]
142 (version 1.1.3) R libraries.

143

144 **Results**

145 Out of 40,273 unique clinical trialists present in the HemOnc KB as of 2023-11-21, 37,420
146 (92.9%) had a resolvable first name and were thus eligible for gender determination. This group was
147 sourced from 7,473 clinical trial manuscripts published between 1947-2023. After excluding trialists with
148 gender not yet determined (n=4,360, 11.7%), those with a determined unknown gender (n=78, 0.2%), and
149 those with a determined gender but initial-only first names (n=14, <0.1%), the final analysis set included
150 32,968 trialists with predetermined binary gender. Of the 32,968 trialists, 11,398 (34.6%) were designated
151 as women. There were 7849 unique names after normalizing first initial/middle name combinations to
152 only include a middle name. The remainder of names were shared by more than one individual. Michael
153 was the most common name, with 473 (1.4%) occurrences. Only 1,899 (24.2%) of names occurred more
154 than twice.

155 Of 25,240 trialists with a known site affiliation, 24,930 (98.8%) were affiliated with sites in a
156 single country and were assigned to the country of their affiliated institution when querying Genderize
157 and Gender API with nationalities. When excluding clinical trialists without a recorded country of origin,
158 the number of trialists and unique names was 24,930 and 6,756, respectively. The final analysis set
159 included trialists from 87 countries, the most abundant being the US with 9,485 (38%) affiliated trialists.
160 There were 7,569 first name-country combinations that occurred only once. The most common first-
161 name-country combination was David-US with 201 (0.8%) instances. Only 1,760 (7.1%) of first name-
162 country combinations appeared more than twice. The 100 most common trialist name-country
163 combinations are presented in **Supplementary Table 1**.

164

165 **Gender prediction accuracy when country of origin was not supplied (baseline case)**

166 The overall accuracy of Genderize when predicting gender for the full dataset without supplying
167 country was 96.6% with 2.3% incorrect gender predictions and 1.1% of names yielding no prediction
168 (**Table 1**). Similarly, the overall accuracy of Gender API was 96.1% with 2.7% incorrect gender
169 predictions and 1.1% of names resulting in no prediction. The accuracy of the gender R package's
170 predictions was lower, with 79.8% and 85.7% accuracy with the IPUMS and SSA methods, respectively.
171 Names of men were misgendered as women less than 3% of the time for all gender prediction tools
172 (**Table 1**). Names of women were misgendered over 3% of the time for all services except the gender R
173 package when using SSA data as a reference. The difference in the percent of correct gender predictions
174 between Genderize and Gender API was significant in favor of Genderize ($p < 0.001$). Likewise, the
175 accuracy difference between the gender R package methods were also significant ($p < 0.001$), in favor of
176 the SSA method. Gender API demonstrated higher gender prediction accuracy when two-part names were
177 delimited with a space: the percent of correctly inferred genders rose from 96.1% to 96.3%.

178

179

180

181 **Table 1: Accuracy of Gender Predictions on 32,968 Included Trialists**

Method ^a	Correct, n (%)	Incorrect, n (%)	No Predictions, n (%)	Men Incorrectly Gendered as Women, n (%) ^b	Women Incorrectly Gendered as Men, n (%) ^b
Genderize	31,857/32,968 (96.6%)	763/32,968 (2.3%)	348/32,968 (1.1%)	401/21,324 (1.9%)	362/11,296 (3.2%)
Gender API	31,690/32,968 (96.1%)	899/32,968 (2.7%)	379/32,968 (1.1%)	393/21,320 (1.8%)	506/11,269 (4.5%)
gender (IPUMS)	26,294/32,968 (79.8%)	1366/32,968 (4.1%)	5308/32,968 (16.1%)	508/17,941 (2.8%)	858/9719 (8.8%)
gender (SSA)	28,266/32,968 (85.7%)	590/32,968 (1.8%)	4112/32,968 (12.5%)	489/18,595 (2.6%)	101/10,261 (1%)

182 ^aTwo-part first names were appended together without a delimiting character.

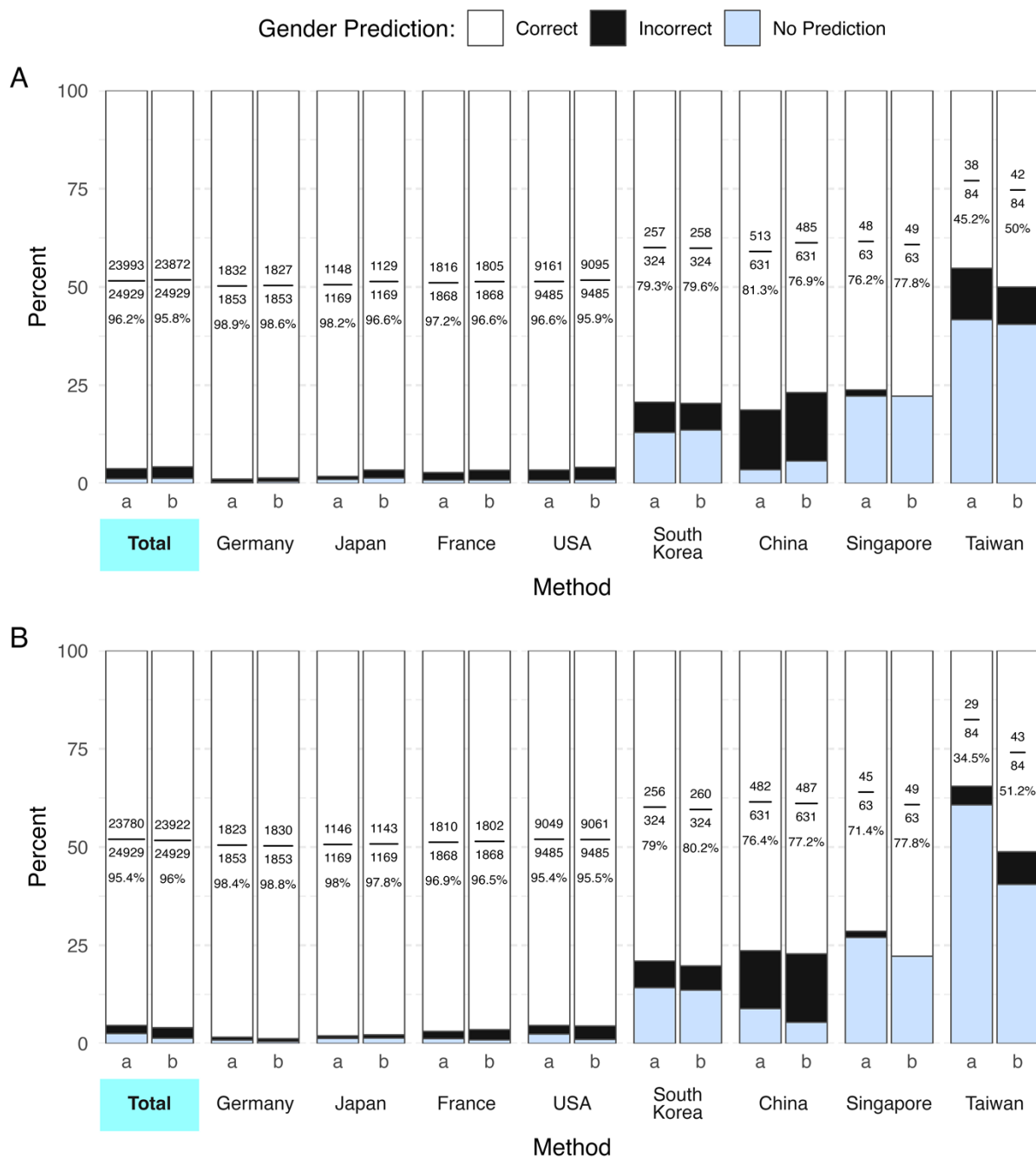
183 ^bDenominators are not consistent across rows because names that did not return a gender prediction for a
184 given service were excluded.

185
186 After restricting Genderize’s predictions to trialists affiliated with a single country, the percentage
187 of correct, incorrect, and missing predictions were 96.2%, 2.6%, and 1.2% respectively (**Fig 1A**).
188 Genderize achieved the highest accuracy when evaluating first names from German authors, and the
189 lowest accuracy when evaluating names from South Korean, Chinese, Singaporean, and Taiwanese
190 authors. When evaluating the same 24,929 clinical trialists with Gender API, the percentage of correct,
191 incorrect, and missing predictions were 95.8%, 3%, and 1.3% respectively. Gender API also had high
192 accuracy when predicting the gender of German authors, and the lowest accuracy when evaluating names
193 from South Korean, Chinese, Singaporean, and Taiwanese authors. The difference in accuracy between
194 Genderize and Gender API is significant ($p < 0.001$), in favor of Genderize.

195
196 **Gender prediction accuracy when country of origin was supplied to the API**

197 The gender prediction accuracies when countries of origin were supplied to Genderize and
198 Gender API are visualized in **Fig 1B**. Supplying the countries of origin alongside first names in the API
199 call decreased the percentage of correct gender predictions when using Genderize from 96.2% to 95.4%,

200 while also reducing the percentage of incorrect predictions from 2.6% to 2.1%. Conversely, including
201 countries of origin increased the ratio of correct gender predictions of Gender API from 95.8% to 96%
202 and decreased incorrect predictions from 3% to 2.7%. Supplying countries also increased the percentage
203 of names with no gender prediction for Genderize from 1.2% to 2.5%, while Gender API remained
204 constant at 1.3%. The difference in accuracy between Genderize and Gender API was significant in favor
205 of Gender API ($p < 0.001$).



206

207 **Fig 1: Accuracy of Gender Predictions.** Panel A shows the gender prediction accuracies when countries

208 are not included in the API call. Panel B shows the results when countries are included in the API call.

209 The top 4 countries with the most trialists and all Asian countries are plotted. Method a is genderize and

210 method b is Gender API. Each bar is labeled with the fraction and percentage of correct gender

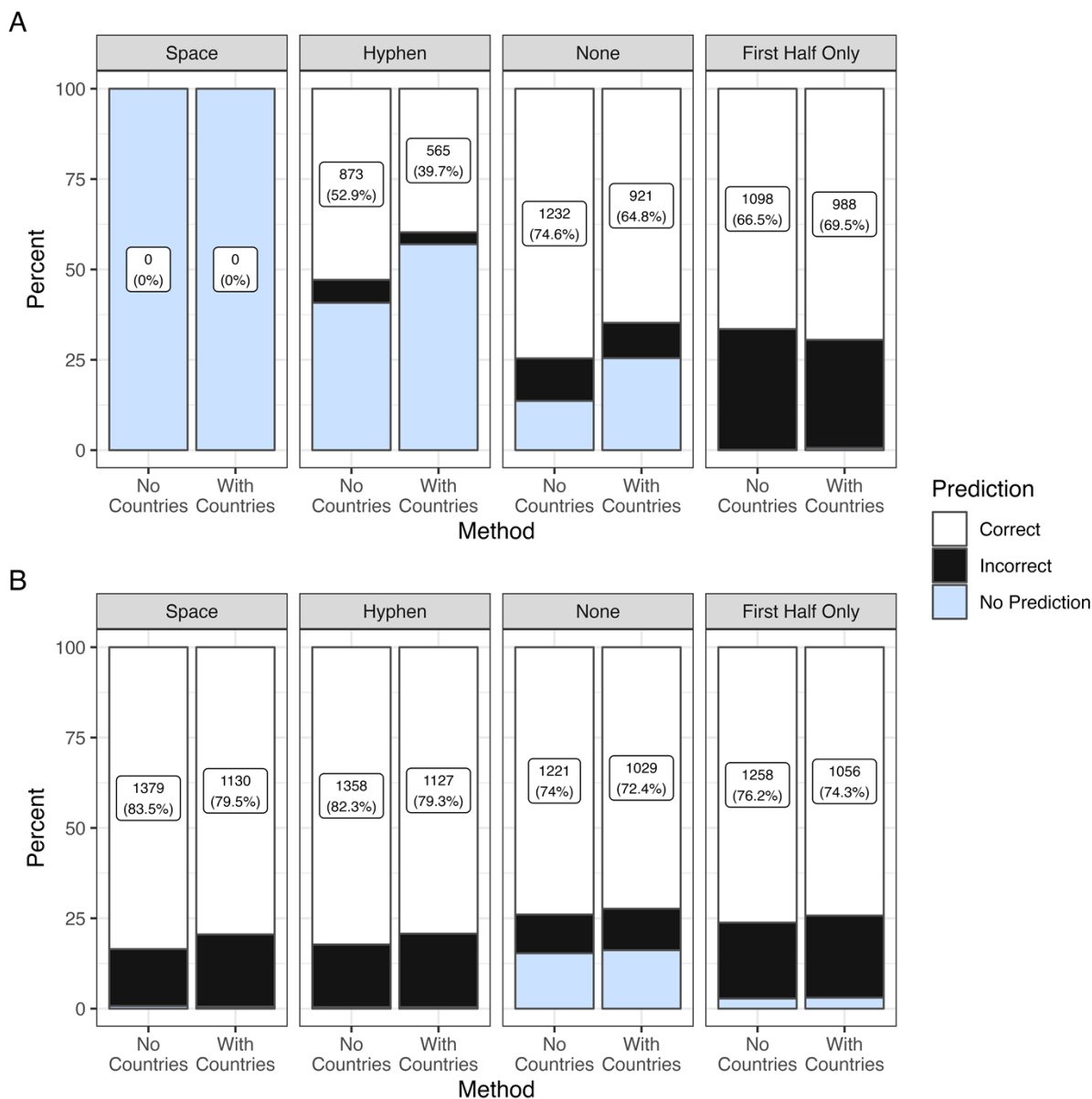
211 predictions. Two-part first names were appended together without a delimiting character.

212 **Gender Prediction accuracy when using different characters to delimit two-part forenames**

213 Gender prediction accuracy when evaluating two-part names was higher when countries were not
214 included in the API call in all contexts except when calling Genderize with the first half of a two-part
215 name, e.g., Jean-Pierre as jean. Genderize was most accurate (76.4%) when no character was used to
216 delimit two-part names, e.g., Jean-Pierre represented as jeanpierre (**Fig 2**). Genderize provided zero
217 predictions for two-part first names delimited with a space. In contrast, Gender API achieved the highest
218 gender prediction accuracy when delimiting two-part names with a space (83.5%). Gender prediction
219 accuracy for two-part names was worse than for one-part names when countries were not included in the
220 API call and two-part names were separated without a delimiter: OR 0.07 (95% CI 0.06-0.08) for
221 Genderize and OR 0.08 (95% CI 0.07-0.09) for Gender API, respectively.

222 The accuracy of Genderize and Gender API were evaluated for statistical significance by
223 comparing the percent of correct gender predictions between delimiter categories. The difference in
224 Gender prediction accuracy between Genderize and Gender API when evaluating two-part names without
225 a delimiting character and including countries in the API call was not significant. All other comparisons
226 between Genderize and Gender API were significant in favor of Gender API ($p < 0.001$).

227



228
 229 **Fig 2: Accuracy of Gender Predictions Based on Delimiter between Two-Part Names.** Panel A is
 230 Genderize and Panel B is Gender API. Plot facets correspond to the type of delimiter separating two-part
 231 names. Stacked bars correspond to correct, incorrect, and no predictions respectively. Bars are labeled
 232 with the count and percent of correct gender predictions.

233
 234
 235

236 **Gender prediction accuracy by API-reported confidence thresholds**

237 There was high agreement overall between gender prediction services and the gold standard
238 labeled dataset (**Fig 3**). Genderize reported over 50% confidence in gender predictions for 32,573 (98.8%)
239 trialists. Similarly, Gender API reported over 50% confidence for 32,587 (98.8%) trialists. Gender API
240 demonstrated a correlation of 0.91 between its reported confidence and actual accuracy, compared to
241 Genderize's correlation of 0.82. The Brier scores for Gender API and Genderize were 0.0077 and 0.0048
242 respectively.

243

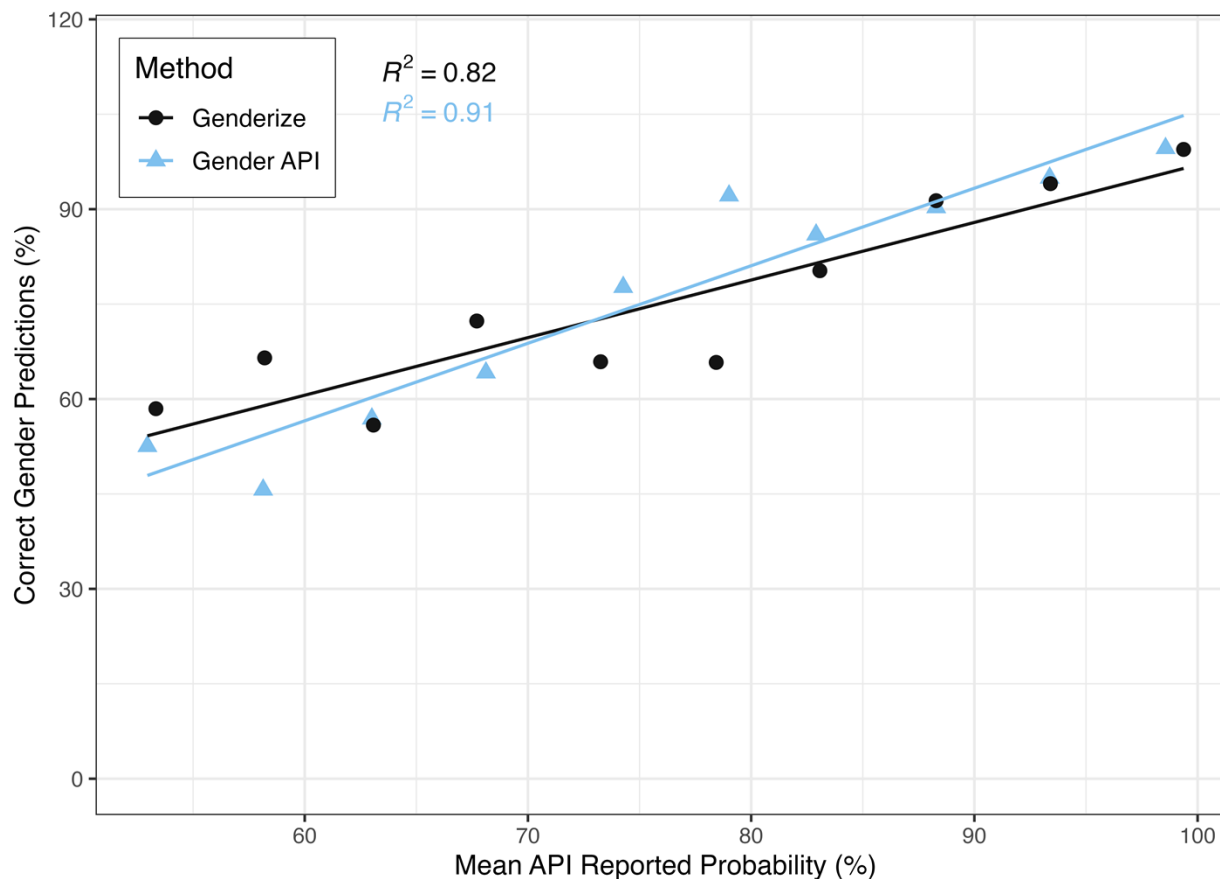
244

245

246

247

248



249

250 **Fig 3: Experimental Name Prediction Accuracy At Different API Probability Cutoffs.** Names with
251 gender predictions were aggregated into the following API reported probability bins: 50%-55%, 55%-
252 60%, 60%-65%, 65%-70%, 70%-75%, 75%-80%, 80%-85%, 85%-90%, 90%-95%, 95%-100%. The API
253 reported probabilities within each bin were averaged and plotted on the x-axis. The experimentally
254 determined gender prediction accuracies for the names in each bin are visualized on the y-axis.

255

256 **Replication of analyses by Santamaria 2018 and Sebo 2021**

257 The original dataset used by Santamaria consisted of 5779 first names with known genders, 34%
258 of whom were women. Only 0.4% of the 5779 had diacritical marks. In addition, 1.1% and 2% of names
259 contain spaces or hyphens, respectively. The original paper reported 80% accuracy using Genderize and
260 87% using Gender API. In our re-analysis, Genderize predicted the correct gender 92.5% of the time.

261 Similarly, Gender API achieved 92.8% accuracy. The difference in accuracy between Genderize and
262 Gender API was not statistically significant.

263 The dataset originally analyzed by Sebo 2021 included 6131 names of whom 50.3% were
264 women. Diacritical marks were present in 6.6% of names. 10.2% of names contained spaces, and 6.6% of
265 names included a hyphen. The original paper reported 81% accuracy using Genderize and 97% with
266 Gender API. In our re-analysis, the accuracy of Genderize and Gender API on these 6131 names were
267 86.2% and 98% respectively. McNemar's test indicated that the differences in accuracy was statistically
268 significant ($p < 0.001$), in favor of Gender API. Gender API was 99.5% accurate when evaluating names
269 with diacritical marks, while Genderize was 71.7% accurate.

270

271 **Cost and Accessibility**

272 Genderize and Gender API provide a graphical user interface, while the gender R package
273 requires programming. Genderize [10] provides 1,000 free predictions per day, whereas Gender API [11]
274 only allows 100 free predictions per month. Gender API currently costs 4.85x more than Genderize for a
275 monthly subscription that provides 100,000 predictions.

276

277 **Discussion**

278 Genderize and Gender API both demonstrated over 95% overall accuracy on our gold-standard
279 dataset of cancer clinical trialists. Genderize was slightly more accurate than Gender API when countries
280 were not included in the API call. Conversely, Gender API performed slightly better than Gender API
281 when countries were included. For both services, including countries reduced the number of incorrect
282 gender assignments at the cost of increasing the number of names with no predicted gender (**Fig 1,**
283 **Supplementary Table 2, Supplementary Table 3**). The gender R package performed worse than
284 Genderize or Gender API (**Table 1**).

285 Genderize and Gender API differed in how their accuracy was affected by the delimiter
286 separating two-part first names. Genderize was most accurate when two-part first names were appended

287 together without a delimiter (**Fig 2**). In fact, Genderize appeared to be incompatible with two-part first
288 names that were delimited by a space as the service yielded zero correct predictions when evaluating such
289 names. Conversely, Gender API performed best when two-part first names were delimited with a space.
290 The slightly higher overall gender prediction accuracy attained by Genderize compared to Gender API is
291 partially an artifact of our decision to append two-part names without a delimiter in the baseline
292 comparison, since Gender API performed best when two-part names were delimited with a space.

293 A commonality between this analysis and several previous studies was the lower prediction
294 accuracy of Genderize and Gender API when evaluating Asian names, with the exception of Japanese
295 names [15,19]. The higher accuracy achieved for both services in our re-analysis of Santamaria’s dataset
296 indicates that both services have improved since 2015, although Genderize improved by a larger margin.
297 Gender API outperformed Genderize when re-analyzing Sebo’s dataset largely because Gender API
298 handled two-part names that were delimited by spaces as well as names with diacritical marks. In fact, a
299 follow up study [14] by the same author recommended removing diacritical marks and modifying two-
300 part names to improve the accuracy of Genderize.

301 This study’s results should be interpreted with certain caveats in mind. We did not filter out
302 recurring first names during this analysis because the count of names in real-world datasets like ours tends
303 to follow a long-tail distribution [29]. The process for determining the “gold-standard” gender of each of
304 the trialists relied on inference from available information. Affiliation data was missing for a substantial
305 subset of authors, mostly due to the older practice of MEDLINE including only the affiliation of the first
306 author; a substantial number of high-profile oncology journals (e.g., the *Journal of Clinical Oncology* and
307 *Blood*) did not include clear 1:1 mappings for author-to-affiliation for a period of time; this issue affects
308 at least 320 (4%) of manuscripts in the HemOnc KB. A substantial subset of recently added authors to the
309 HemOnc KB have not had their gender determined yet (13.7%), and this subset has some important
310 differences from the set of determined genders. Most notably, the undetermined subset has many more
311 Asian hyphenated names (43.8% vs 3.8%) and authors with a country of affiliation including South
312 Korea, China, Singapore, and/or Taiwan (45.6% vs 3.42%). It is thus likely that our results represent a

313 “best-case scenario” and that automated gender mapping will become increasingly difficult as cancer
314 clinical trials are increasingly conducted in the Asia-Pacific region [30,31]. Additionally, a researcher’s
315 nationality in our data set does not always reflect the cultural origin of their first name as some
316 researchers immigrated to the country of their academic affiliation.

317 It is important to note that Genderize, Gender API, and the gender R package assume a gender
318 binary. However, a recent survey [32] found that 1.6% of U.S. adults identify as transgender or
319 nonbinary. With new algorithmic advancements such as Genderize and Gender API, it is imperative that
320 inclusivity is incorporated. Going forward, tools that infer gender based on name should be trained on
321 data that include transgender and nonbinary people, and they should include the option to predict an
322 individual as non-binary or transgender. Gender prediction is not simply a binary classification problem.
323 Transgender individuals will likely make up a small percentage of the dataset of names, which could
324 make obtaining a correct prediction for these individuals very challenging. Yet, by not incorporating
325 them, we are excluding countless people from this algorithm and ensuring the prediction of their gender
326 to have a 0% accuracy.

327 Both Genderize and Gender API demonstrated high gender prediction accuracy with Western
328 names that were highly normalized without middle or last names or diacritical marks. The cost per name
329 evaluated with Genderize is also several times cheaper than Gender API. However, Genderize loses
330 accuracy compared to Gender API when name formatting becomes less consistent. The SSA and IPUMS
331 methods of the gender R package were less accurate but are open-source alternatives. The results from
332 this study provide a new benchmark for gender inference tools.

333

334 **Alexander D. VanHelene:** Conceptualization, Methodology, Software, Validation, Formal analysis,
335 Investigation, Data Curation, Writing - Original Draft, Writing - Review & Editing, Visualization.

336 **Ishaani Khatri:** Conceptualization, Writing - Original Draft, Writing - Review & Editing. **C. Beau**

337 **Hilton:** Formal analysis, Writing - Review & Editing. **Sanjay Mishra:** Methodology, Formal analysis,

338 Investigation, Data Curation, Visualization, Writing - Review & Editing, Funding acquisition. **Ece D.**

339 **Gamsiz Uzun:** Visualization, Writing - Review & Editing. **Jeremy L. Warner:** Conceptualization,
340 Methodology, Validation, Formal analysis, Investigation, Resources, Data Curation, Writing - Original
341 Draft, Writing - Review & Editing, Visualization, Supervision, Funding acquisition.

342

343 **Acknowledgements**

344 We would like to acknowledge the efforts of the editorial board of HemOnc.org.

345

346

347 **References**

- 348 1. Chatterjee P, Werner RM. Gender Disparity in Citations in High-Impact Journal Articles. *JAMA*
349 *Netw Open*. 2021;4: e2114509. doi:10.1001/jamanetworkopen.2021.14509
- 350 2. Murphy M, Callander JK, Dohan D, Grandis JR. Women’s Experiences of Promotion and Tenure in
351 Academic Medicine and Potential Implications for Gender Disparities in Career Advancement: A
352 Qualitative Analysis. *JAMA Netw Open*. 2021;4: e2125843.
353 doi:10.1001/jamanetworkopen.2021.25843
- 354 3. Dymanus KA, Butaney M, Magee DE, Hird AE, Luckenbaugh AN, Ma MW, et al. Assessment of
355 gender representation in clinical trials leading to FDA approval for oncology therapeutics between
356 2014 and 2019: A systematic review-based cohort study. *Cancer*. 2021;127: 3156–3162.
357 doi:10.1002/cncr.33533
- 358 4. Wais K. Gender Prediction Methods Based on First Names with genderizeR. *R J*. 2016;8/1: 17–37.
- 359 5. Cevik M, Haque SA, Manne-Goehler J, Kuppalli K, Sax PE, Majumder MS, et al. Gender disparities
360 in coronavirus disease 2019 clinical trial leadership. *Clin Microbiol Infect*. 2021;27: 1007–1010.
361 doi:10.1016/j.cmi.2020.12.025
- 362 6. Topaz CM, Sen S. Gender Representation on Journal Editorial Boards in the Mathematical Sciences.
363 Danforth CM, editor. *PLOS ONE*. 2016;11: e0161357. doi:10.1371/journal.pone.0161357
- 364 7. Nielsen MW, Andersen JP, Schiebinger L, Schneider JW. One and a half million medical papers
365 reveal a link between author gender and attention to gender and sex analysis. *Nat Hum Behav*.
366 2017;1: 791–796. doi:10.1038/s41562-017-0235-x
- 367 8. Sebo P, Clair C. Are female authors under-represented in primary healthcare and general internal
368 medicine journals? *Br J Gen Pract*. 2021;71: 302.1-302. doi:10.3399/bjgp21X716249
- 369 9. Szymkowiak M. Genderizing fisheries: Assessing over thirty years of women’s participation in
370 Alaska fisheries. *Mar Policy*. 2020;115: 103846. doi:10.1016/j.marpol.2020.103846
- 371 10. Genderize Documentation. In: Genderize [Internet]. [cited 2 Jan 2024]. Available:
372 <https://genderize.io/>
- 373 11. Gender API - Determines the gender of a first name. [cited 2 Jan 2024]. Available: [https://gender-
374 api.com/?utm_source=adwords&utm_medium=cpc&utm_campaign=ga3&price-
375 set=OTG&gad_source=1&gclid=Cj0KCQiAhc-
376 sBhCEARIsAOVwHuQF7HdhUmWIWZfU851GAvR14ziBWaEc6tDDR_XKmG6I904GgzaqYr4a
377 AoCfEALw_wcB](https://gender-api.com/?utm_source=adwords&utm_medium=cpc&utm_campaign=ga3&price-set=OTG&gad_source=1&gclid=Cj0KCQiAhc-sBhCEARIsAOVwHuQF7HdhUmWIWZfU851GAvR14ziBWaEc6tDDR_XKmG6I904GgzaqYr4aAoCfEALw_wcB)
- 378 12. Mullen L. gender: Predict Gender from Names Using Historical Data. 2021. Available:
379 <https://github.com/lmullen/gender>
- 380 13. Sebo P. How accurate are gender detection tools in predicting the gender for Chinese names? A study
381 with 20,000 given names in Pinyin format. *J Med Libr Assoc*. 2021;110.
382 doi:10.5195/jmla.2022.1289

- 383 14. Sebo P. Using genderize.io to infer the gender of first names: how to improve the accuracy of the
384 inference. *J Med Libr Assoc.* 2021;109. doi:10.5195/jmla.2021.1252
- 385 15. Santamaría L, Mihaljević H. Comparison and benchmark of name-to-gender inference services. *PeerJ*
386 *Comput Sci.* 2018;4: e156. doi:10.7717/peerj-cs.156
- 387 16. Warner JL, Cowan AJ, Hall AC, Yang PC. HemOnc.org: A Collaborative Online Knowledge
388 Platform for Oncology Professionals. *J Oncol Pract.* 2015;11: e336–e350.
389 doi:10.1200/JOP.2014.001511
- 390 17. Heidari S, Babor TF, De Castro P, Tort S, Curno M. Sex and Gender Equity in Research: rationale
391 for the SAGER guidelines and recommended use. *Res Integr Peer Rev.* 2016;1: 2.
392 doi:10.1186/s41073-016-0007-6
- 393 18. CIHR Institute Of Gender And Health. What a difference sex and gender make : a gender, sex and
394 health research casebook. 2012 [cited 18 Jan 2024]. doi:10.14288/1.0132684
- 395 19. Sebo P. Performance of gender detection tools: a comparative study of name-to-gender inference
396 services. *J Med Libr Assoc.* 2021;109. doi:10.5195/jmla.2021.1185
- 397 20. Mihaljevic H, Santamaria L. Evaluation of name-based gender inference methods. *GenderGapSTEM-*
398 *PublicationAnalysis*; 2023. Available: [https://github.com/GenderGapSTEM-](https://github.com/GenderGapSTEM-PublicationAnalysis/name_gender_inference)
399 [PublicationAnalysis/name_gender_inference](https://github.com/GenderGapSTEM-PublicationAnalysis/name_gender_inference)
- 400 21. Sebo P. Performance of gender detection tools: a comparative study of name-to-gender inference
401 services. 2021 [cited 2 Jan 2024]. doi:10.17605/OSF.IO/KR2MX
- 402 22. R Core Team. R: A Language and Environment for Statistical Computing. R Foundation for
403 Statistical Computing; Available: <https://www.R-project.org/>
- 404 23. Wickham H, Averick M, Bryan J, Chang W, McGowan L, François R, et al. Welcome to the
405 Tidyverse. *J Open Source Softw.* 2019;4: 1686. doi:10.21105/joss.01686
- 406 24. Wickham H, Miller E, Smith D. haven: Import and Export “SPSS”, “Stata” and “SAS” Files. 2023.
407 Available: <https://CRAN.R-project.org/package=haven>
- 408 25. Wickham H, Bryan J. readxl: Read Excel Files. 2023. Available: [https://CRAN.R-](https://CRAN.R-project.org/package=readxl)
409 [project.org/package=readxl](https://CRAN.R-project.org/package=readxl)
- 410 26. Wickham H. testthat: Get Started with Testing. 2011. Available: [https://journal.r-](https://journal.r-project.org/archive/2011-1/RJournal_2011-1_Wickham.pdf)
411 [project.org/archive/2011-1/RJournal_2011-1_Wickham.pdf](https://journal.r-project.org/archive/2011-1/RJournal_2011-1_Wickham.pdf)
- 412 27. Aphalo P. ggpmisc: Miscellaneous Extensions to “ggplot2.” 2023. Available: [https://CRAN.R-](https://CRAN.R-project.org/package=ggpmisc)
413 [project.org/package=ggpmisc](https://CRAN.R-project.org/package=ggpmisc)
- 414 28. Pedersen T. patchwork: The Composer of Plots. 2023. Available: [https://CRAN.R-](https://CRAN.R-project.org/package=patchwork)
415 [project.org/package=patchwork](https://CRAN.R-project.org/package=patchwork)
- 416 29. Clauset A, Shalizi CR, Newman MEJ. Power-Law Distributions in Empirical Data. *SIAM Rev.*
417 2009;51: 661–703. doi:10.1137/070710111

- 418 30. Akiki V, Troussard X, Metges J, Devos P. Global trends in oncology research: A mixed-methods
419 study of publications and clinical trials from 2010 to 2019. *Cancer Rep.* 2023;6: e1650.
420 doi:10.1002/cnr2.1650
- 421 31. Terada M, Nakamura K, Matsuda T, Okuma HS, Sudo K, Yusof A, et al. A new era of the Asian
422 clinical research network: a report from the ATLAS international symposium. *Jpn J Clin Oncol.*
423 2023;53: 619–628. doi:10.1093/jjco/hyad033
- 424 32. Minkin AB Juliana Menasce Horowitz, Kim Parker and Rachel. The Experiences, Challenges and
425 Hopes of Transgender and Nonbinary U.S. Adults. In: Pew Research Center’s Social &
426 Demographic Trends Project [Internet]. 7 Jun 2022 [cited 2 Jan 2024]. Available:
427 [https://www.pewresearch.org/social-trends/2022/06/07/the-experiences-challenges-and-hopes-of-](https://www.pewresearch.org/social-trends/2022/06/07/the-experiences-challenges-and-hopes-of-transgender-and-nonbinary-u-s-adults/)
428 [transgender-and-nonbinary-u-s-adults/](https://www.pewresearch.org/social-trends/2022/06/07/the-experiences-challenges-and-hopes-of-transgender-and-nonbinary-u-s-adults/)
- 429
- 430
- 431