

1 Use of machine learning methods to understand discussions of female
2 genital mutilation/cutting on social media

3
4 Gray Babbs,^{1#a¶} Sarah E. Weber,^{1#b¶*} Salma M. Abdalla,¹ Nina Cesare,^{2&} Elaine O. Nsoesie^{3&}

5
6 ¹ Department of Epidemiology, Boston University School of Public Health, Boston,
7 Massachusetts, United States of America

8 ² Biostatistics and Epidemiology Data Analytics Center, Boston University School of Public
9 Health, Boston, Massachusetts, United States of America

10 ³ Department of Global Health, Boston University School of Public Health, Boston,
11 Massachusetts, United States of America

12 ^{#a} Current Address: Department of Health Services, Policy, and Practice, Brown University
13 School of Public Health, Providence, Rhode Island, United States of America

14 ^{#b} Current Address: Section of Infectious Diseases, Boston Medical Center, Boston,
15 Massachusetts, United States of America

16
17
18

19 ¶ These authors contributed equally to this work

20 & These authors also contributed equally to this work

21
22 * Corresponding author
23 sarah.weber@bmc.org

24
25
26

27 **Abstract**

28 Female genital mutilation/cutting (FGM/C) describes several procedures that involve injury to
29 the vulva or vagina for nontherapeutic reasons. Though at least 200 million women and girls
30 living in 30 countries have undergone FGM/C, there is a paucity of studies focused on public
31 perception of FGM/C. We used machine learning methods to characterize discussion of FGM/C
32 on Twitter in English from 2015 to 2020. Twitter has emerged in recent years as a source for
33 seeking and sharing health information and misinformation. We extracted text metadata from
34 user profiles to characterize the individuals and locations involved in conversations about
35 FGM/C. We extracted major discussion themes from posts using correlated topic modeling.
36 Finally, we extracted features from posts and applied random forest models to predict user
37 engagement. The volume of tweets addressing FGM/C remained fairly stable across years.
38 Conversation was mostly concentrated among the United States and United Kingdom through
39 2017, but shifted to Nigeria and Kenya in 2020. Some of the discussion topics associated with
40 FGM/C across years included Islam, International Day of Zero Tolerance, current news stories,
41 education, activism, male circumcision, human rights, and feminism. Tweet length and follower
42 count were consistently strong predictors of engagement. Our findings suggest that (1)
43 discussion about FGM/C has not evolved significantly over time, (2) the majority of the
44 conversation about FGM/C on English-speaking Twitter is advocating for an end to the practice,
45 (3) supporters of Donald Trump make up a substantial voice in the conversation about FGM/C,
46 and (4) understanding the nuances in how people across cultures refer to and discuss FGM/C
47 could be important for the design of public health communication and intervention.

48

49

50 **Author Summary**

51 Female genital mutilation/cutting (FGM/C) refers to procedures that involve injury to the vulva
52 and vagina for nontherapeutic reasons. At least 200 million women and girls living in 30
53 countries have undergone FGM/C. In many countries, there is not public perception research on
54 FGM/C. When research is conducted, survey respondents may not answer truthfully about their
55 opinions. We used Twitter data from 2015 to 2020 to better understand how the public thinks
56 about FGM/C. We found that the conversation has generally stayed the same in this time
57 period. Early in the study, users discussing FGM/C were mostly from the United States and
58 United Kingdom, but later in the study, users were mostly from Nigeria and Kenya. Many people
59 posted about stopping FGM/C, especially on the International Day of Zero Tolerance for Female
60 Genital Mutilation and using the hashtag #endfgm. Some people used FGM/C to justify
61 Islamophobia and connected it with other practices like honor killings and acid attacks. We hope
62 that this research can be used to inform communication about FGM/C in different countries to
63 improve understanding about the practice and, design culturally effective campaigns to end
64 FGM/C.

65

66 **Introduction**

67 Female genital mutilation/cutting (FGM/C) is an umbrella term for procedures that
68 involve injury to the vulva and vagina for nontherapeutic reasons.¹ At least 200 million women
69 and girls living in 30 countries have undergone FGM/C.² Rates of FGM/C are generally
70 decreasing, but the rates by which they are decreasing vary by country and region.^{3,4} FGM/C is
71 most prevalent in Eastern, Northeastern, and Western Africa and also occurs among migrants

72 and refugees worldwide.¹ FGM/C is practiced across religions and is not solely associated with
73 any religious group.^{5,6} FGM/C is illegal in many countries, including nearly all of the countries
74 where it is most commonly practiced.⁷

75 FGM/C is considered a violation of human rights law and a pressing public health issue.⁸
76 It is associated with a number of negative short-term health outcomes, including hemorrhage,
77 shock, and infection, as well as long-term physical complications, including chronic infections,
78 painful urination, sexual health challenges, and obstetric complications.⁹ Studies have also
79 pointed to a relationship between FGM/C and adverse mental health outcomes, including post-
80 traumatic stress disorder (PTSD), anxiety and depression.¹⁰

81 Negative public discourse on FGM/C can perpetuate stigma for women and girls in
82 diasporic communities that practice (or are perceived to practice) FGM/C.¹¹ A qualitative study
83 reported that people who have undergone FGM/C feel pressure to lie about their FGM/C status
84 to avoid stigmatization or social exclusion in their new cultural contexts.¹² Other studies note
85 that cut women in the diaspora report a perceived sense of “feeling different” from their
86 peers,^{13,14} above and beyond differences due to immigration status and race/ethnicity alone.¹⁵
87 This stigma causes many women to completely avoid discussion about FGM/C.¹⁴

88 Analysis of social media discussion offers the opportunity to assess sentiments,
89 knowledge and attitudes towards FGM/C globally. Social media platforms offer users the ability
90 to connect with global communities. Their perceived anonymity can bring traditionally private
91 conversations into the public sphere, serving as valuable resources for understanding
92 perceptions of stigmatized issues.^{16–18} Social media data are uniquely positioned to track public
93 conversations over time, particularly for sensitive topics that may be difficult to capture via
94 surveys such as, miscarriage, preterm birth¹⁹ and circumcision.²⁰ Moreover, social media offers
95 the opportunity to analyze how issues are framed within a conversation space, as they contain

96 longitudinal information, and user-to-user replies. However, few studies to date have used social
97 media data to analyze FGM/C discussions.²¹

98 We aim to characterize the substance of conversations about FGM/C on Twitter
99 between 2015 and 2020 and assess any changes over time. Our analysis focuses on
100 conversation prevalence, geographic distribution, framing and themes, frequently used terms,
101 characteristics of users engaging in the conversation, and factors influencing engagement. This
102 study is the first of its scale to describe and analyze social media conversations about FGM/Cs.
103 Our findings can aid medical professionals, public health practitioners, governmental agencies,
104 and advocacy organizations seeking to leverage social media to inform FGM/C-related
105 communications and interventions.

106 **Materials and methods**

107 **User characteristics**

108 To understand user volume and account type, we compared the number of users who
109 tweeted about FGM/C each year, the number of users with verified accounts, and the proportion
110 of users each year who were newly engaged. We defined new engagement as users who
111 tweeted about FGM/C for the first time for each year during the study period.

112 We examined top terms found in user descriptions by normalizing, stemming, and
113 removing all stop words from user description text and calculating the frequency of each word
114 used. We extracted the top 15 words used each year. Overall, a total of 25 words were included
115 in at least one year's top 15 most frequently used words. Finally, we calculated the proportion of
116 user descriptions each year containing each word and compared over time.

117 Using extracted country names, we calculated the proportion of total users each year
118 who identified each country in their user location, and we evaluated changes in the most
119 frequently identified countries across years.

120 **Characterizing FGM/C conversations over time**

121 We measured the number of tweets, retweets and likes in each year of the study period.
122 We measured tweet volume on International Day of Zero Tolerance for FGM by assessing
123 number of tweets on February 6 of each year in the study period.

124 We also evaluated the use of FGM/C terminology by determining the frequencies of the
125 terms “female circumcision,” “cutting,” “mutilation,” “excision,” and “alteration.” Presence of
126 terms in a tweet were not mutually exclusive.

127 Top words in tweet text were found using the same method as the user description. We
128 extracted the top 15 words used each year and calculated the proportion of tweets each year
129 containing each word. Overall, 34 words were included.

130 To characterize conversations about FGM/C, we used Correlated Topic Modelling (CTM)
131 to identify major topics in our corpus.^{22,23} CTM is an extension of latent Dirichlet allocation (LDA)
132 topic modeling, which assumes that documents (tweets) are generated from a finite group of
133 topics. CTM is distinguished, however, by the recognition that topics within a document may be
134 correlated with one another. We ran CTM models separately for each year of analysis. Topics
135 were inferred by the research team based on groupings of terms that were most commonly
136 found in and most unique to a topic. Topics were cross-validated by referencing the text of
137 tweets containing each topic’s unique terms. Ten topics were selected for each year, after
138 exploratory analysis. Reported example tweets have been modified slightly to preserve user

139 anonymity, according to best practices.²⁴ Tweet topics that were specific to an event or
140 particular year were synthesized using only tweets from the year the topic was identified.

141 **Understanding user engagement**

142 We analyzed tweet and user characteristics associated with greater user engagement
143 using two classifiers – support vector machine (SVM) and random forest models, which have
144 been used in similar studies.^{25,26} We fit models first on all of the years in the data. We then fit
145 models on each year's subset, using the number of retweets to represent user engagement.
146 Number of retweets was classified into two categories: high retweets were defined as the year's
147 90th retweet percentile or above, and low retweets were a random matched sample of tweets
148 with zero retweets. We chose retweet count as the representation of user engagement because
149 it had little missingness compared to like count, favorite count, reply count, and quote count. We
150 split all user engagement models into train and test sets using a 70/30 split ratio.

151 We optimized the random forest model with varying numbers of trees and inclusion of
152 predictors. Account level variables in the model included number of accounts followed, number
153 of accounts following, number of tweets tweeted, and account verification. Tweet level variables
154 included tweeting on a high volume date (defined as a top 11 date for the dataset), tweet length,
155 and inclusion of key FGM/C terminology in the tweet (“alteration,” “circumcision,” “cutting,”
156 “excision,” “mutilation,” “fgm”). We compared accuracy between the SVM and random forest
157 models and ran separate models for each of the six study years. All analyses were completed in
158 R Version 4.0.2.

159 **Ethics Statement**

160 The protocol was approved by the Boston University Medical Campus Institutional Review
161 Board (Protocol Number: H-41567) and determined exempt. Consent was not obtained from
162 participants because all data were collected from publicly available sources. The study posed
163 no risks to individuals or individual privacy.

164 Results

165 User characteristics

166 During the study period from 2015 to 2020, 338,845 unique users tweeted about FGM/C.
167 The most users were engaged in conversations in 2017 and 2018, and the least users were
168 engaged in 2016 (Table 1). The number of users decreased steadily from 2017 to 2020. Users
169 were similar from 2015-2016; only 23.4% of users in 2016 had not previously tweeted about
170 FGM/C in 2015. However, in 2017, 83.7% of users joined FGM/C discussions for the first time
171 during the study period and the percentage of newly engaged users remained high from 2018-
172 2020. A similar percentage of users each year were verified, excluding a spike in verified users
173 in 2016.

174

175 **Table 1. Number of users engaged in conversations around FGM/C over the study period**
176 **and frequency of new and verified users within each year**

| Year | Total Users | New Users N (%) | Verified Users N (%) |
|------|-------------|--------------------|-------------------------|
| 2015 | 82,034 | — | 3,759 (4.6) |
| 2016 | 26,291 | 6,160 (23.4) | 1,840 (7.0) |
| 2017 | 95,162 | 79,669 (83.7) | 3,942 (4.1) |
| 2018 | 90,368 | 63,300 (70.0) | 3,622 (4.0) |
| 2019 | 86,978 | 57,711 (66.4) | 3,176 (3.7) |
| 2020 | 72,957 | 49,971 (68.5) | 2,483 (3.4) |

177

178 The most common terms in user descriptions were “views,” “news,” “health,” and “love”
179 (Fig 1). Other prevalent terms included “world,” “writer,” “social,” “rights,” “life,” and “Trump.”
180 Notably, “MAGA” and “conservative” were also among the top user description terms. When
181 comparing user description terms from year-to-year, there were no considerable changes in
182 term prevalence over time.

183 **Fig 1. Most Prevalent Terms in User Descriptions in 2020**

184 We linked 100,267 (30%) of our 338,845 unique users to countries based on user
185 descriptions. The conversation was concentrated among six countries (United States, United
186 Kingdom, Nigeria, Kenya, India and Australia), though relative proportions changed over time
187 (Fig 2). In 2015 and 2016, the highest proportion of users were from the UK (20% and 21%
188 respectively), followed by the United States, Nigeria, Kenya and Australia. From 2017 to 2019,
189 users were heavily concentrated in the United States (between 19% and 24% of all users during
190 this time). In 2020, the countries with the highest proportion of users were Kenya (15%) and
191 Nigeria (14%), followed by the USA (10%), UK (10%), and India (7%).

192 **Fig 2. User Geographic Distribution in 2015, 2017 and 2020**

193 **Characterizing FGM/C conversations over time**

194 The overall volume of tweets addressing FGM/C remained moderately stable between
195 2015 and 2020 (Fig 3). We observed the greatest number of tweets (291,187) in 2017. This
196 represents a two-fold jump in the number of tweets over the preceding year. Likes and retweets
197 increased steadily between 2016 and 2018, after which likes continued to increase and retweets
198 dropped.

199 **Fig 3: Twitter Engagement by Year.** Tweet count is represented by gray bars, like count is
 200 represented by the blue line, and retweet count is represented by the pink line.

201 In each of the study years, the highest volume of FGM/C specific tweets occurred on
 202 February 6, International Day of Zero Tolerance for Female Genital Mutilation, representing
 203 4.8% of the entire dataset. The average daily tweet volume on this date was 16.7 times higher
 204 than the daily average.

205 The terms that were used to refer to FGM/C remained fairly stable across the years
 206 studied. In addition to female genital mutilation or FGM, cutting (4.3%), female circumcision
 207 (2.2%), excision (0.1%), and alteration (<0.1%) were commonly used terms for this practice.
 208 The words most commonly found in tweet text from 2015-2020 included “endfgm,”
 209 “circumcision,” “child,” and “news’ (Fig 4).’ Locations were often referenced, including the UK,
 210 Nigeria, Michigan, Kenya and Egypt. “Muslim,” “Islam,” and “Sharia” were also among the most
 211 commonly used words in the tweet text.

212 **Fig 4: Most Prevalent Terms Used in Tweets in 2020**

213 Despite similarity in commonly referenced terms, topics identified using topic modeling
 214 varied across years (Table 2). Overall, Islam and the International Day of Zero Tolerance for
 215 Female Genital Mutilation were the most common topics identified. Both topics were present in
 216 all six years of analysis. Activism was the next most persistent topic, occurring in each year after
 217 2015. Comparisons between FGM/C and male circumcision arose in 2017 and were present for
 218 every subsequent year. Conversations regarding the legality of FGM/C were identified in four
 219 years (2015, 2016, 2018, and 2019).

220 **Table 2. Topics Identified within FGM/C Tweets by Year**

| | 2015 | 2016 | 2017 | 2018 | 2019 | 2020 | Tweet Example |
|-------|------|------|------|------|------|------|---|
| Islam | | | | | | | “I’m personally really ashamed that fgm |

| | | | | | | | |
|---|--|------------------|--|--|--|--|---|
| | | | | | | | is practiced in so much of the muslim world. It's so cruel..." |
| International Day of Zero Tolerance for FGM | | | | | | | "Today is International Day of Zero Tolerance. What is our government doing to stop FGM? #endFGM" |
| Current News Stories | | | | | | | "Michigan Doctor Arrested for Female Genital Mutilation" |
| Education | | | | | | | "Everyone needs to be educated on #FGM to protect vulnerable girls." |
| Legal | | | | | | | "Sierra Leone is close to outlawing FGM. Huge win." |
| Cultural Aspects | | | | | | | "FGM is a cultural practice- not a religious one. Culture can change, and the culture of FGM needs to be ended immediately..." |
| Medical Risks | | | | | | | "I saw the lacerations, the fistulas, the difficulties giving birth after. FGM is the most harmful thing that can happen to a girl." |
| Activism | | | | | | | "New challenges in the fight to ban FGM" |
| (Male) Circumcision | | | | | | | "Why does no one talk about male genital mutilation?" |
| "Evils of the Other" | | | | | | | "I don't consider FGM/honour killings/forced marriages /Sharia/Halal/terrorism etc. to be part of British culture. It's sad that you do." |
| Feminism | | | | | | | "FGM is why we need feminism" |
| Human Rights | | | | | | | "FGM isn't a female problem. It's a human rights problem." |
| Additional Identified Topics | | Violence; UNICEF | | | | | Violence: "Worldwide, 1 in 3 women experience sexual or physical violence. 200 million have undergone FGM. #endviolenceagainstwomen" UNICEF: "Rates of FGM are decreasing in Africa but disturbingly increasing in Indonesia." (Note: this topic corresponded with the 2016 release of a |

| | | | | | | | |
|--|--|--|--|--|--|--|--|
| | | | | | | | UNICEF report about FGM/C.) ² |
|--|--|--|--|--|--|--|--|

221 A shaded box indicates that a topic was found in the topic model for a given year subset.

222 Example tweets have been modified slightly to preserve user anonymity.

223 In 2017, 2018, 2019, and 2020, a topic emerged that put FGM/C in a category with other
224 human rights issues specific to women and girls (“acid attacks,” “honor killing,” “gang rape,” and
225 “child marriage”). In Table 2, this topic is referred to as “evils of the other.” Tweets associated
226 with these topics are often anti-immigrant and Islamophobic. Education around FGM/C was a
227 topic present from 2015-2017, but it was not present from 2018-2020. This topic includes terms
228 that promote FGM/C awareness such as “speak,” “question” and “believe.” FGM/C as a
229 culturally significant practice was a topic identified across three years: 2015, 2016 and 2018.

230 Other themes identified across multiple years included medical risks (2015, 2019),
231 human rights (2018, 2020), and feminism (2017, 2020). A number of topics could be linked to
232 specific events, such as the implementation of stricter laws against FGM/C in Eritrea (2015),²⁷
233 The Gambia banning FGM/C (2015),²⁸ Egypt’s parliament approving stringent penalties for
234 FGM/C (2016),²⁹ a doctor in Michigan, USA being charged with performing FGM/C (2017),^{30,31}
235 and Sudan outlawing FGM/C (2020).^{32,33} In 2020, COVID-19 emerged as a theme that was
236 correlated with FGM/C discussions.

237 **User engagement**

238 On average, original tweets received 3.86 retweets (standard deviation (SD)=53.88) and
239 all tweets received an average of 3.49 favorites (SD=93.90). Using the full dataset, the random
240 forest model performed better for predicting user engagement, with an 86.3% accuracy (F1:
241 0.8621) compared to 73.0% in the SVM model (F1: 0.7440). Results from the random forest
242 model suggest that factors predicting engagement were stable across time (Fig 5). Number of

243 followers was consistently the first or second most important predictor, with the highest mean
244 decrease in Gini score in 2015 and 2017 (S1 Table). Longer tweet length was the most
245 important predictor in 2016 and 2018-2020. Variables such as the following count, number of
246 total tweets, and whether the user had been verified, were also important predictors of whether
247 a tweet would garner a high number of retweets. The presence of common words for FGM/C in
248 the tweet text were of relatively little importance in the model. The model accuracy generally
249 increased in later years.

250 **Fig 5. Mean Decrease Gini Score by Tweet and User Characteristics**

251 Discussion

252 In this study, we aimed to characterize discussions of FGM/C on the social media
253 platform, Twitter, from 2015 to 2020. In general, we found that discussion about FGM/C has not
254 changed significantly over time. However, there were differences in the number of newly
255 engaged users over the years. In 2017, 83.7% of users joined FGM/C discussions for the first
256 time. While tweet count steadily decreased after 2017, retweet and like counts remained well
257 above pre-2017 levels. This may suggest that although fewer original ideas about FGM/C were
258 circulating post-2017, users were more familiar with the topic and more likely to retweet other
259 voices. Major countries where conversation was concentrated remained consistent; however,
260 the country with the highest volume of users shifted from the UK to the United States in 2017,
261 until Nigeria and Kenya became the top countries in 2020.

262 We identify substantial increases in conversation related to five news stories in the study
263 period: stricter laws against FGM/C in Eritrea (2015), The Gambia banning FGM/C (2015),
264 Egypt's parliament approving stringent penalties for FGM/C (2016), a doctor in Michigan, USA
265 charged with performing FGM/C (2017), and Sudan outlawing FGM/C (2020). Islam as a topic

266 was identified across all years and “Muslim,” “Islam,” and “Sharia” were among the most
267 common words in tweet text. These findings suggest that users often associated FGM/C with
268 Islam, despite FGM/C occurring across religions.

269 Our results suggest that a substantial contingent of the conversation happening about
270 FGM/C on Twitter (in English) is advocating for an end to FGM/C. The hashtag “endfgm” was
271 the most common unique word identified in tweet text. Each year in the study period, we saw a
272 substantial surge in tweets for International Day of Zero Tolerance for FGM, a 16.7-fold increase
273 over the daily average. This suggests that observing specific days when FGM/C is recognized
274 globally can increase awareness since individuals are likely to engage in conversation and
275 advocacy on such days. Activism, education, feminism, and human rights were persistent topics
276 that arose in the data across years.

277 Another significant portion of the conversation used FGM/C as a rhetorical tool to
278 associate Islam with what we have termed “evils of the other,” a new topic which arose in 2017
279 and connected FGM/C to other practices such as child marriage, acid attacks, female
280 infanticide, and gang rape. This shift coincides with a 2017 increase in the proportion of users in
281 the US engaging in FGM/C conversations. “MAGA,” “Trump,” and “conservative” were three of
282 the most common words in Twitter user profiles of people tweeting about FGM/C, suggesting a
283 strong voice from conservative Trump supporters in the FGM/C conversation. FGM/C sentiment
284 that connects the practice closely with other phenomena that are considered “of the other” in the
285 West may further stigmatize people who have undergone FGM/C by painting “narratives of
286 victimhood”.³⁴

287 These findings have substantive implications for public health communication and
288 practice. First, on English speaking Twitter, the term FGM was most frequently used to describe
289 the practice. Communication about FGM/C should contain the term female genital mutilation or

290 FGM to align with common usage. Second, endfgm was the single most common word
291 identified in our dataset. Use of this hashtag can connect public health communication with the
292 greater FGM/C advocacy conversation. Third, given that user verification was the best predictor
293 of engagement, verified accounts have a particularly large role to play in engaging the public in
294 education about FGM/C. In addition, these results suggest that the conversation about FGM/C
295 is increasingly politicized in English speaking contexts, particularly in the UK and US, as seen
296 by the strong conservative voice in the conversation. Providers and advocates who serve
297 patients who have undergone FGM/C should practice cultural humility and be informed about
298 the potential stigmatization of FGM/C in the current sociopolitical moment. Finally, in public
299 health communication about FGM/C, specific emphasis should be placed on supporting
300 survivors and uplifting their voices.

301 One major limitation of this work is that it is restricted to conversations happening about
302 FGM/C on Twitter. Twitter users are not representative of the global public, and we have limited
303 information on demographic characteristics of users. Conversations offline or on different social
304 media sites may vary from the findings here. Further, the data are not definitively representative
305 of all conversations or users. Though tweets were collected from users around the world, we
306 only analyze discourse among English language tweets. Tweets about FGM/C were identified
307 based on presence of certain terms, so we may overlook conversations about FGM/C that do
308 not use these terms, particularly terms for FGM/C in languages other than English. Lastly, topic
309 models were run separately for each year of data. As such, the models are not directly
310 comparable and may contain slightly different identified terms.

311 Despite these limitations, this study uses a large sample size to characterize
312 conversations around this important public health topic. With over a million original tweets
313 posted between 2015 and 2020, these data span a wide range of populations and
314 conversations. It is the first study to our knowledge to evaluate FGM/C related social media

315 discussions using machine learning methods. Though some surveys have been conducted
316 about public perception about FGM/C, many of these surveys used direct questioning. Past
317 studies have indicated that direct questioning about FGM/C, particularly among people who
318 have been engaged in anti-FGM/C related programming, may lead to misreporting due to social
319 desirability bias.³⁵ Social media platforms such as Twitter have substantial potential to study
320 sensitive issues like FGM/C, as they do not employ direct questioning, but rather observe
321 ongoing conversations.

322 Future studies should explore what proportion of social media conversations about
323 FGM/C are from people immediately affected by the practice. This work may also address how
324 discussion of FGM/C is impacted by events (e.g., media, news coverage, or celebrity posts).
325 There is room to explore in greater qualitative detail how discourse varies by country, and
326 whether negative discourse can be linked, on an aggregate level, to stress or negative health
327 outcomes. Given the usage of such platforms for sharing health information and misinformation,
328 it is important to understand how discussion about topics such as FGM/C are impacted by
329 misinformation and political discourse. With additional research, public health practitioners, non-
330 governmental organizations, and governments can educate and design targeted interventions to
331 reach communities and countries where the practice is ongoing.

332 These analyses add to a growing body of work that uses social media conversations to
333 characterize stigmatized or misunderstood issues. Using Twitter data allows us to keep a finger
334 on the pulse of public discourse to understand perceptions, develop public health
335 communications, and hone more effective interventions.³⁶

References

1. World Health Organization. Female genital mutilation. Published January 21, 2022. Accessed March 25, 2022. <https://www.who.int/news-room/fact-sheets/detail/female-genital-mutilation>
2. United Nations Children's Fund. Female Genital Mutilation/Cutting: A Global Concern. Published online February 2016. Accessed March 27, 2022. <https://data.unicef.org/resources/female-genital-mutilationcutting-global-concern/>
3. Sussman N. After School in Brooklyn, West African Girls Share Memories of a Painful Ritual. *The New York Times*. <https://www.nytimes.com/2011/04/26/nyregion/brooklyn-girls-from-west-africa-recall-genital-cutting.html>. Published April 26, 2011. Accessed April 12, 2022.
4. Touiba N. Female Genital Mutilation. In: *Women's Rights, Human Rights: International Feminist Perspectives*. Routledge; 1995:224-237.
5. Hayford SR, Trinitapoli J. Religious Differences in Female Genital Cutting: A Case Study from Burkina Faso. *J Sci Study Relig*. 2011;50(2):252-271. Accessed May 19, 2022. <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4064295/>
6. Setegn T, Lakew Y, Deribe K. Geographic Variation and Factors Associated with Female Genital Mutilation among Reproductive Age Women in Ethiopia: A National Population Based Survey. *PLOS ONE*. 2016;11(1):e0145329. doi:10.1371/journal.pone.0145329
7. Leye E, Deblonde J, García-Añón J, et al. An analysis of the implementation of laws with regard to female genital mutilation in Europe. *Crime Law Soc Change*. 2007;47(1):1-31. doi:10.1007/s10611-007-9055-7
8. World Health Organization, Department of Reproductive Health and Research. *Eliminating Female Genital Mutilation: An Interagency Statement*.; 2008. Accessed March 27, 2022. <http://www.who.int/reproductivehealth/publications/fgm/9789241596442/en/>

9. Berg RC, Underland V, Odgaard-Jensen J, Fretheim A, Vist GE. Effects of female genital cutting on physical health outcomes: a systematic review and meta-analysis. *BMJ Open*. 2014;4(11):e006316. doi:10.1136/bmjopen-2014-006316
10. Abdalla SM, Galea S. Is female genital mutilation/cutting associated with adverse mental health consequences? A systematic review of the evidence. *BMJ Glob Health*. 2019;4(4):e001553. doi:10.1136/bmjgh-2019-001553
11. O'Neill S, Pallitto C. The Consequences of Female Genital Mutilation on Psycho-Social Well-Being: A Systematic Review of Qualitative Research. *Qual Health Res*. 2021;31(9):1738-1750. doi:10.1177/10497323211001862
12. Ahlberg BM, Jylkäs E, Krantz I. Gendered construction of sexual risks: implications for safer sex among young people in Kenya and Sweden. *Reprod Health Matters*. 2001;9(17):26-36. doi:10.1016/S0968-8080(01)90005-9
13. Kahn S. "You see, one day they cut": The evolution, expression, and consequences of resistance for women who oppose female genital cutting. *J Hum Behav Soc Environ*. 2016;26(7-8):622-635. doi:10.1080/10911359.2016.1238805
14. Parikh N, Saruchera Y, Liao LM. It is a problem and it is not a problem: Dilemmatic talk of the psychological effects of female genital cutting. *J Health Psychol*. 2020;25(12):1917-1929. doi:10.1177/1359105318781904
15. Jordal M, Griffin G, Sigurjonsson H. 'I want what every other woman has': reasons for wanting clitoral reconstructive surgery after female genital cutting – a qualitative study from Sweden. *Cult Health Sex*. 2019;21(6):701-716. doi:10.1080/13691058.2018.1510980
16. Balsamo D, Bajardi P, Salomone A, Schifanella R. Patterns of Routes of Administration and Drug Tampering for Nonmedical Opioid Consumption: Data Mining and Content Analysis of Reddit Discussions. *J Med Internet Res*. 2021;23(1):e21212. doi:10.2196/21212
17. Wekerle C, Vakili N, Stewart SH, Black T. The utility of Twitter as a tool for increasing

- reach of research on sexual violence. *Child Abuse Negl.* 2018;85:220-228.
doi:10.1016/j.chiabu.2018.04.019
18. Chua AYK, Banerjee S. The topic of terrorism on Yahoo! Answers: questions, answers and users' anonymity. *Aslib J Inf Manag.* 2020;72(1):1-16. doi:10.1108/AJIM-08-2019-0204
 19. Cesare N, Oladeji O, Ferryman K, et al. Discussions of miscarriage and preterm births on Twitter. *Paediatr Perinat Epidemiol.* 2020;34(5):544-552. doi:10.1111/ppe.12622
 20. Ucar T, Culpan M, Caskurlu T, Karaman Mİ, Silay MS. The activity and discussion points of #Circumcision through Twitter; a microblogging platform. *Int J Impot Res.* 2018;30(5):249-252. doi:10.1038/s41443-018-0058-y
 21. Wondwossen Fantaye A, Konkle AT. Social media representation of female genital cutting: A YouTube analysis. *Womens Health Lond Engl.* 2020;16:1745506520949732.
doi:10.1177/1745506520949732
 22. Blei DM, Lafferty JD. A correlated topic model of Science. *Ann Appl Stat.* 2007;1(1):17-35.
doi:10.1214/07-AOAS114
 23. Lafferty J, Blei D. Correlated Topic Models. In: *Advances in Neural Information Processing Systems*. Vol 18. MIT Press; 2005. Accessed April 12, 2022.
<https://proceedings.neurips.cc/paper/2005/hash/9e82757e9a1c12cb710ad680db11f6f1-Abstract.html>
 24. Markham A. Fabrication as Ethical Practice. *Inf Commun Soc.* 2012;15(3):334-353.
doi:10.1080/1369118X.2011.641993
 25. Dahllöf M. Automatic prediction of gender, political affiliation, and age in Swedish politicians from the wording of their speeches—A comparative study of classifiability. *Lit Linguist Comput.* 2012;27(2):139-153. doi:10.1093/lit/fqs010
 26. Cesare N, Grant C, Nsoesie E. Detection of User Demographics on Social Media: A Review of Methods and Recommendations for Best Practices. Published online February 6, 2017. <https://www.semanticscholar.org/paper/Detection-of-User-Demographics-on->

Social-Media%3A-A-Cesare-Grant/43b55cc08eba13fb309e77a1033da0ad60bd7a5e

27. 28 Too Many. *Eritrea: The Law and FGM*. Thomas Reuters Foundation; 2018.
[https://www.28toomany.org/static/media/uploads/Law%20Reports/eritrea_law_report_v1_\(july_2018\).pdf](https://www.28toomany.org/static/media/uploads/Law%20Reports/eritrea_law_report_v1_(july_2018).pdf)
28. Lyons K. The Gambia bans female genital mutilation. *The Guardian*.
<https://www.theguardian.com/society/2015/nov/24/the-gambia-bans-female-genital-mutilation>. Published November 24, 2015. Accessed March 3, 2022.
29. Human Rights Watch. Egypt: New Penalties for Female Genital Mutilation. Human Rights Watch News. Published September 9, 2016. Accessed March 3, 2022.
<https://www.hrw.org/news/2016/09/09/egypt-new-penalties-female-genital-mutilation>
30. Kennedy M. Michigan Doctor Charged With Performing Female Genital Mutilation On Girls. *NPR*. <https://www.npr.org/sections/thetwo-way/2017/04/14/523917425/michigan-doctor-charged-with-performing-female-genital-mutilation-on-girls>. Published April 14, 2017. Accessed March 3, 2022.
31. Swaine J. Michigan doctor charged with carrying out female genital mutilation. *The Guardian*. <https://www.theguardian.com/us-news/2017/apr/13/michigan-doctor-charged-female-genital-mutilation-fgm>. Published April 13, 2017. Accessed March 3, 2022.
32. United Nations. Sudan bans female genital mutilation, UNICEF vows to help support new law. UN News. Published May 2, 2020. Accessed March 3, 2022.
<https://news.un.org/en/story/2020/05/1063122>
33. Walsh D. In a Victory for Women in Sudan, Female Genital Mutilation Is Outlawed. *The New York Times*. <https://www.nytimes.com/2020/04/30/world/africa/sudan-outlaws-female-genital-mutilation-.html>. Published April 30, 2020. Accessed March 3, 2022.
34. Momoh C. *Female Genital Mutilation*. CRC Press; 2005.
35. De Cao E, Lutz C. Sensitive Survey Questions: Measuring Attitudes Regarding Female Genital Cutting Through a List Experiment. *Oxf Bull Econ Stat*. 2018;80(5):871-892.

doi:10.1111/obes.12228

36. Sinnenberg L, Buttenheim AM, Padrez K, Mancheno C, Ungar L, Merchant RM. Twitter as a Tool for Health Research: A Systematic Review. *Am J Public Health*. 2017;107(1):e1-e8.

doi:10.2105/AJPH.2016.303512

Supporting information captions

S1 Table. Ranked Importance of Tweet and User Characteristics in Predicting High Retweets in Random Forest Model. The importance is ranked ordinally. The factor listed as “1” is the first most influential factor in engagement, the factor listed as “2” is the second most important, and so on.

medRxiv preprint doi: <https://doi.org/10.1101/2022.07.14.22277640>; this version posted July 15, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. This article is a US Government work. It is not subject to copyright under 17 USC 105 and is also made available for use under a CC0 license.

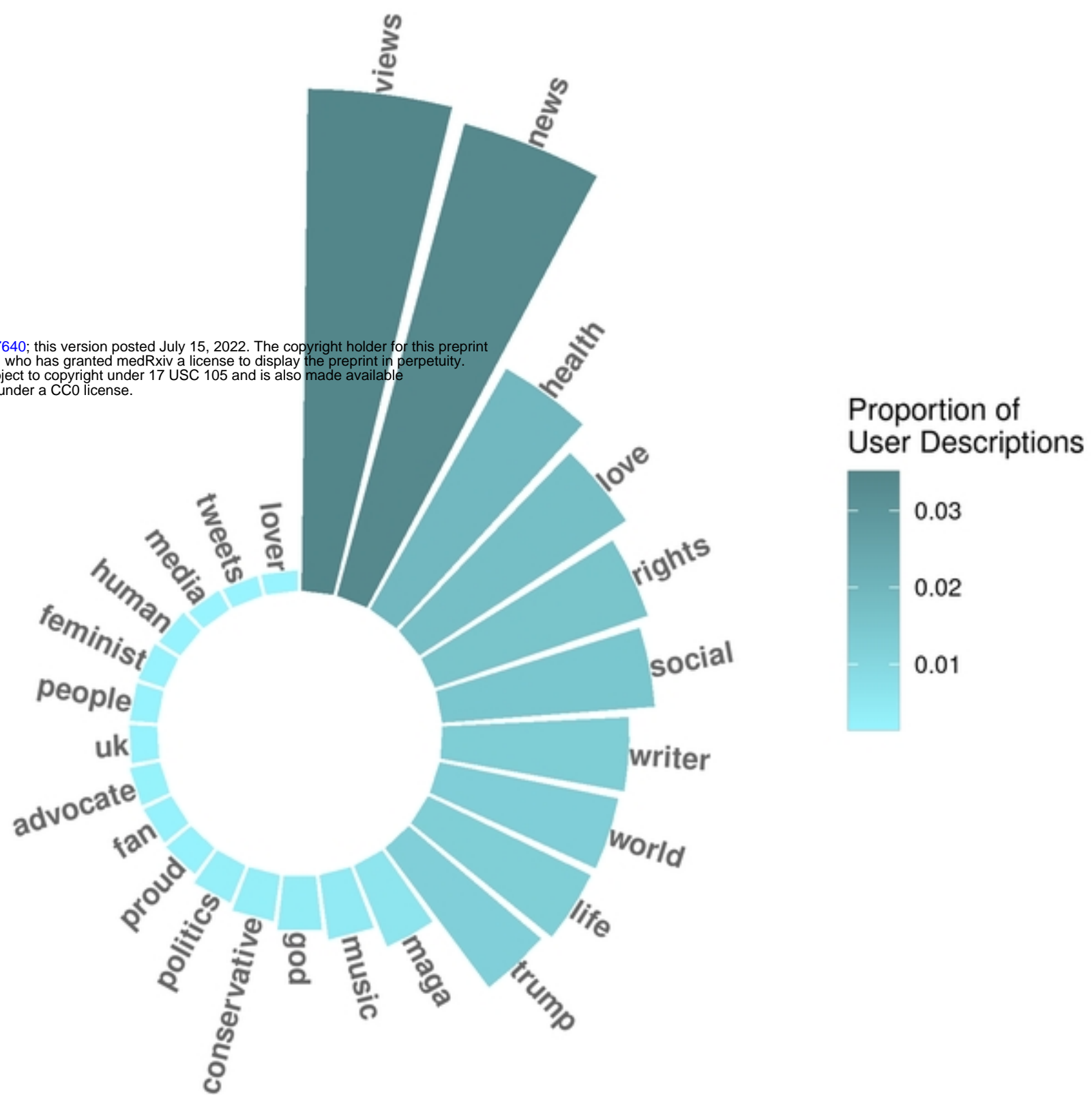


Fig1

Mean Decrease Gini Score by Variable

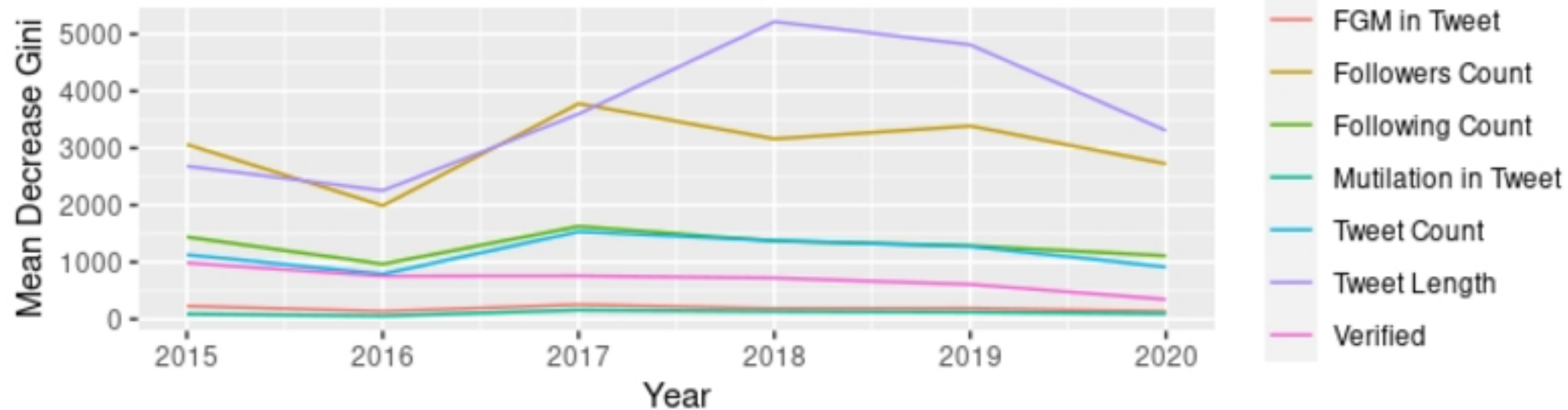


Fig5

Engagement by Year

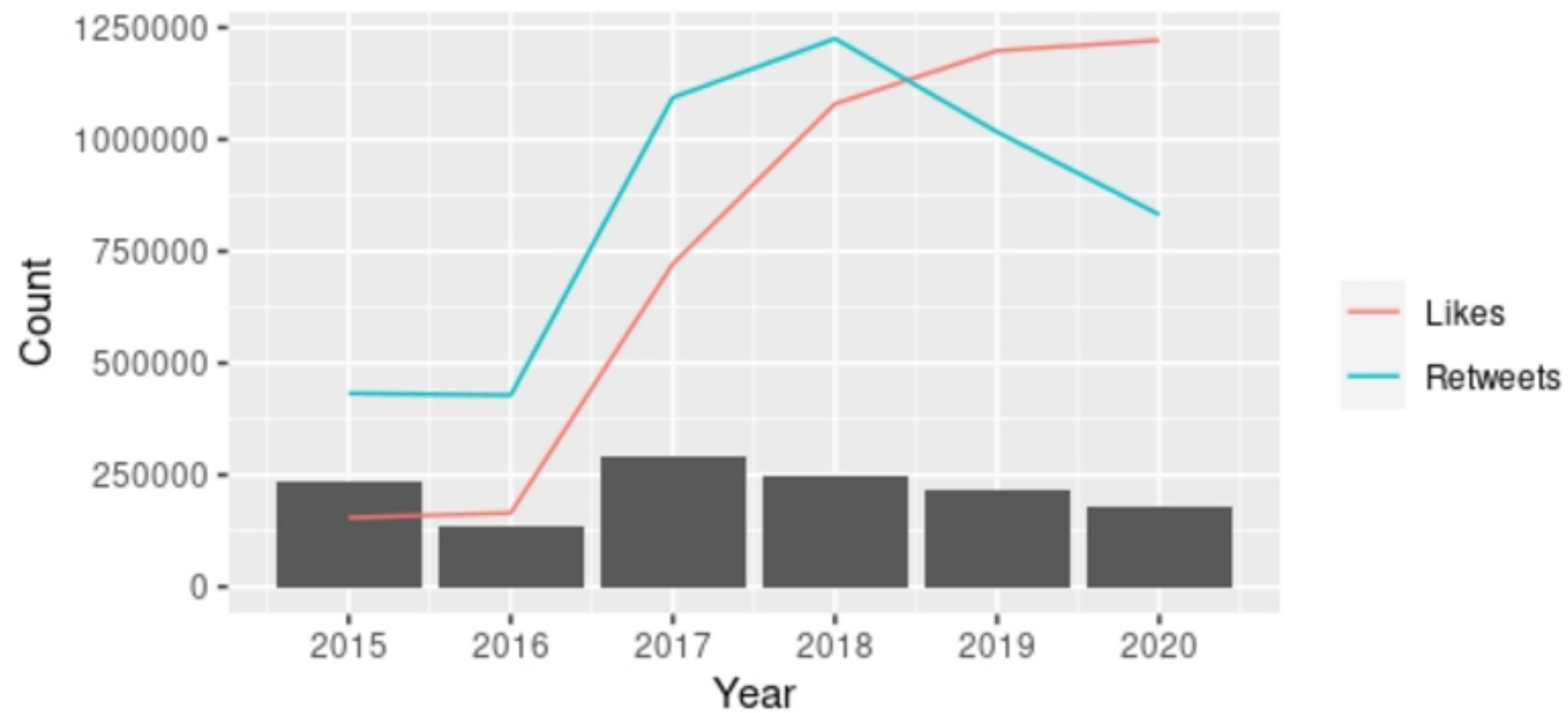


Fig3

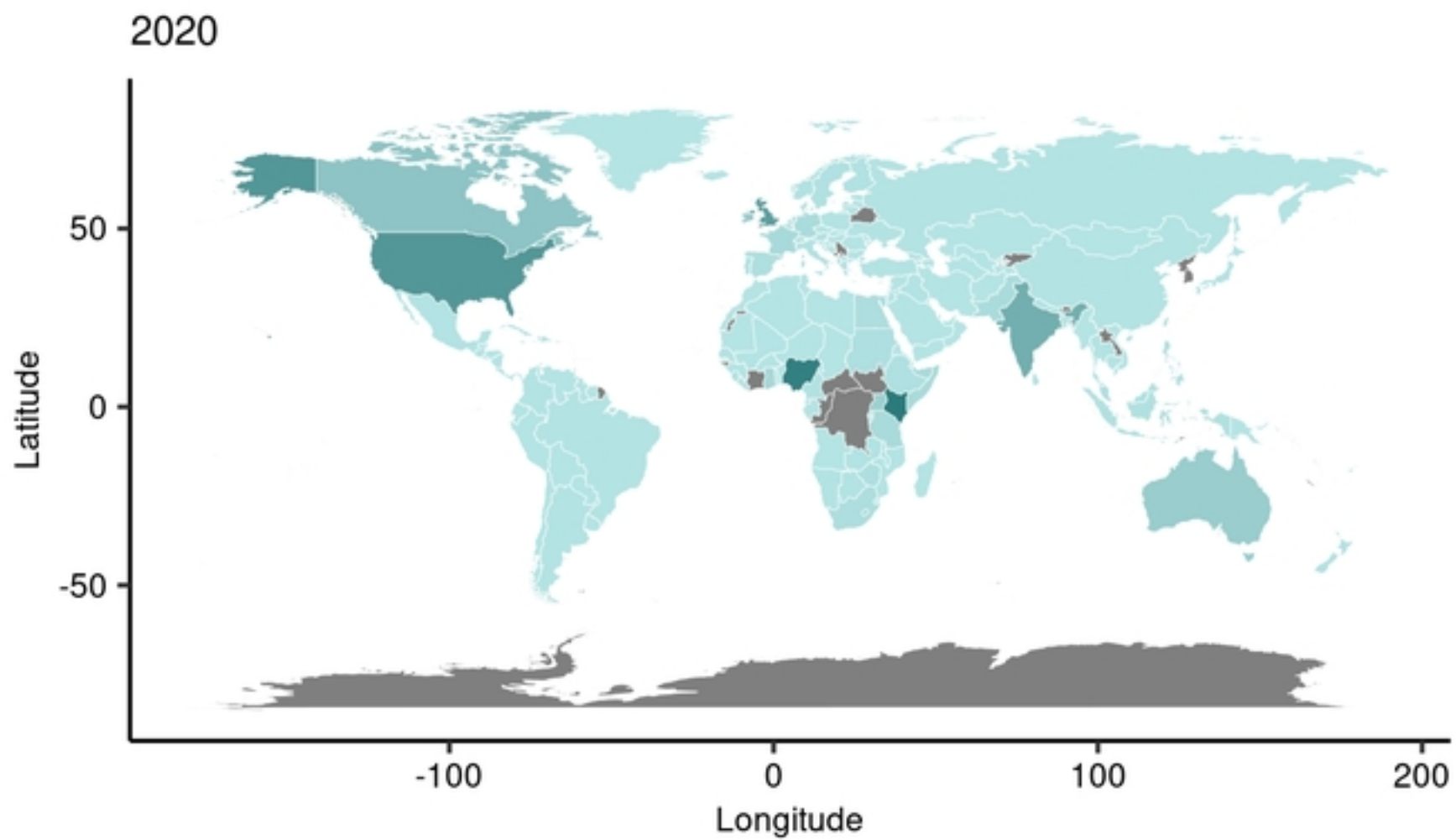
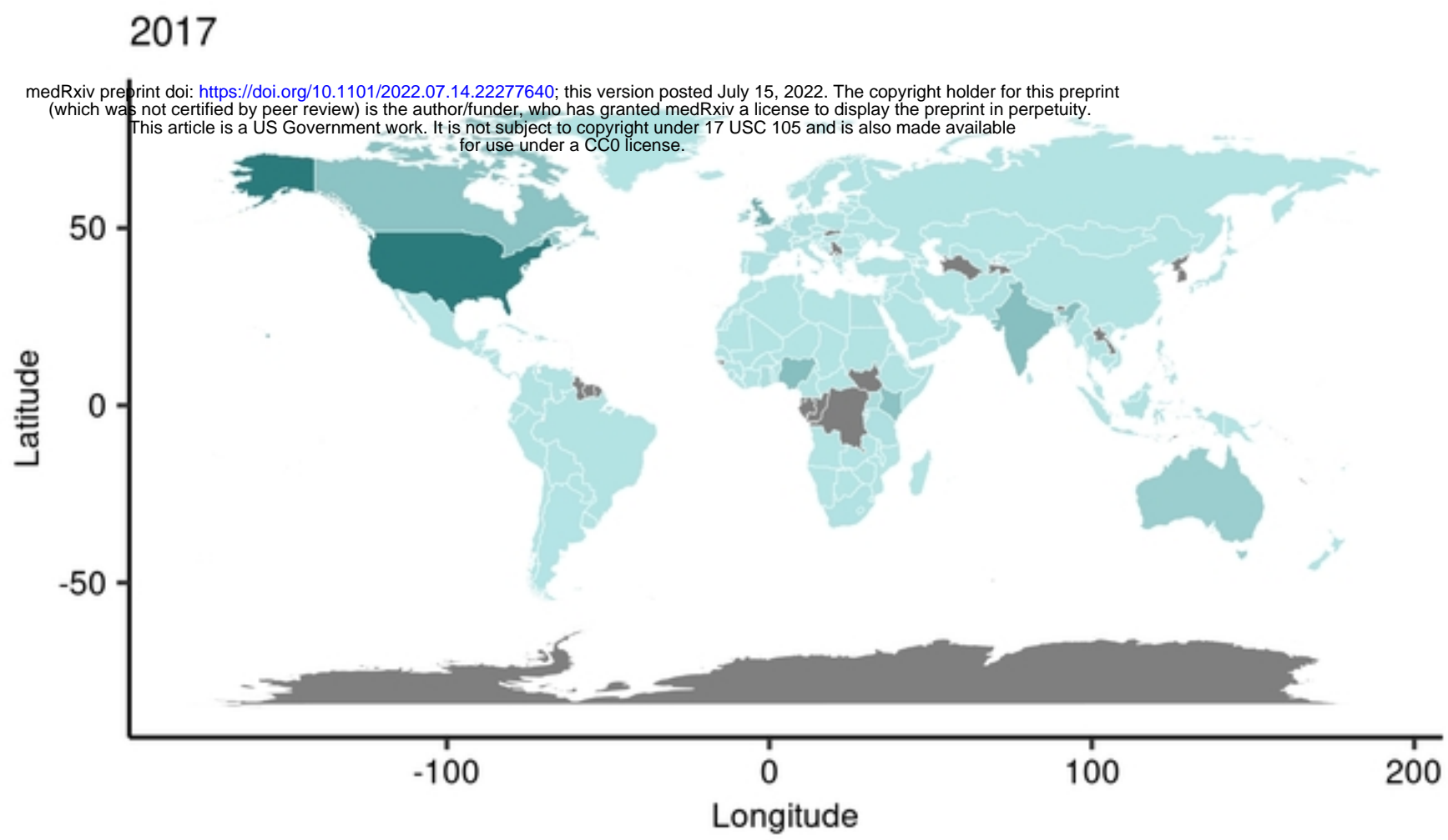
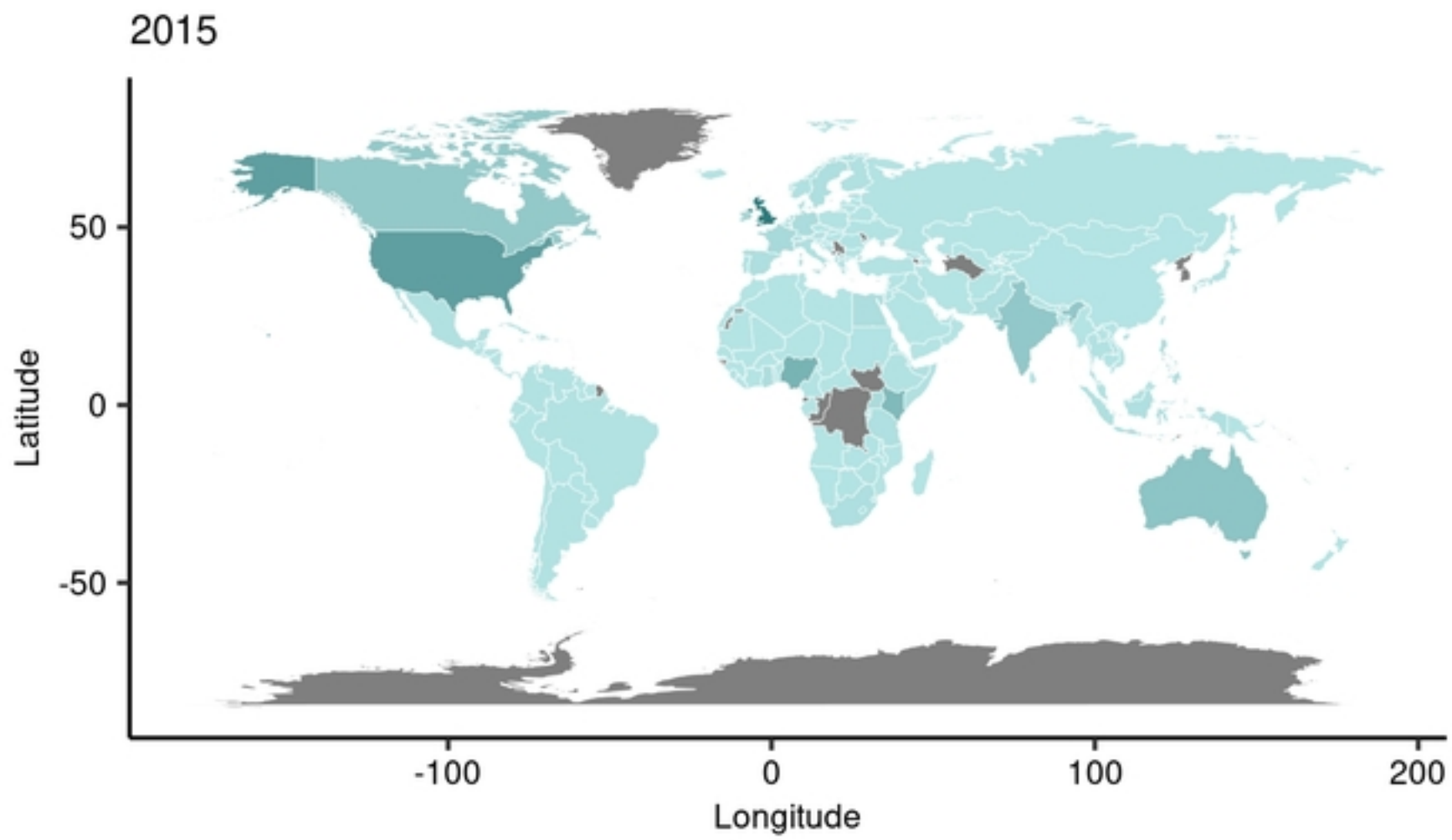


Fig2

medRxiv preprint doi: <https://doi.org/10.1101/2022.07.14.22277640>; this version posted July 15, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. This article is a US Government work. It is not subject to copyright under 17 USC 105 and is also made available for use under a CC0 license.

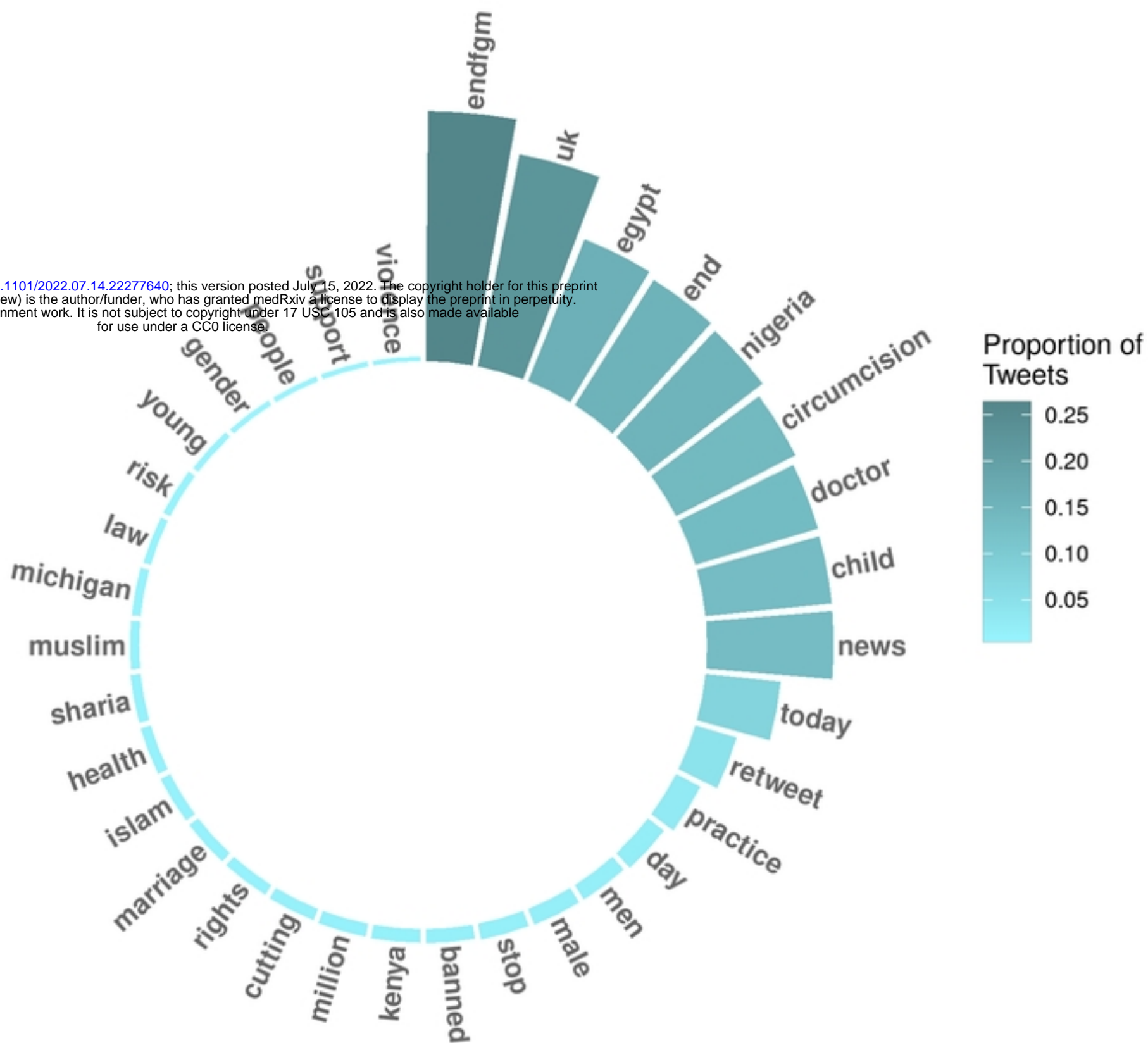


Fig4