The use of Twitter to explore trends in attitudes toward contraceptive methods

by

Allison Merz

Submitted in Partial Fulfillment of the Requirements for the M.D. Degree with Honors in a Special Field at Harvard Medical School

February 10, 2020

Abstract

Background: Unlike other medications, contraceptive methods are often chosen based on the experiences and opinions of individuals' social networks. Though social media, including Twitter, increasingly influences reproductive-age people, discussion of contraception in this setting has yet to be characterized. Natural Language Processing (NLP), a type of machine learning in which computers analyze natural language data, enables this analysis. This study aims to use NLP to explore attitudes toward different contraceptive methods, including both Long- and Short-Acting Reversible Contraception (LARC and SARC), on Twitter since 2006.

Methods: We collected English language tweets mentioning reversible, prescription contraceptive methods with typical-use Pearl Indices of <10 pregnancies per 100 woman-years, including prescription LARC (the intrauterine device (IUD) and the contraceptive implant) and SARC (oral contraceptive pills; the contraceptive patch; the vaginal ring; and the Depo-Provera shot) between March 2006 and December 2019. We used the Amazon Comprehend NLP Sentiment Analysis Application Programming Interface to determine the sentiment of all tweets mentioning a single contraceptive method and evaluated the NLP algorithm's performance based on ten human reviewers' manual sentiment analysis of a random sample of 1000 tweets. All data and code to reproduce this analysis are available at https://github.com/hms-dbmi/contraceptionOnTwitter, and the initial steps can be replicated by launching the code at https://tinyurl.com/cleanTweetsMyBinder.

Results: The number of annual tweets mentioning contraception has increased nearly three hundred-fold since 2007. Out of 838,739 total filtered tweets mentioning at least one contraceptive method, the most commonly tweeted-about method was the IUD (45.9%). LARC methods were mentioned more than SARC methods (58% vs. 42%), and the proportion of LARC-related tweets increased over time. Out of 665,064 tweets mentioning a single contraceptive method, there were nearly twice as many positive tweets about LARC methods compared to SARC methods (19.65% vs. 10.21%, $p<0.05$), though the greatest proportion of all tweets was negative (40.66%). Observed trends in the number and sentiment of tweets about individual contraceptive methods may reflect their historical context including regulation, advertising, availability, and satisfaction, though we did not investigate causal relationships between historical events and tweet volume or content in this analysis.

Implications: Twitter is a potentially valuable source of data for consumer-level discourse regarding contraception and how attitudes toward individual methods have evolved over the past 13 years. Tweets may improve our insight into the perspectives of a traditionally difficult-to-reach population, related to a topic that is often stigmatized. Recognizing the influence of social media on people's lives, and potentially when

considering initiation of a contraceptive method, this and other social media platforms may allow clinicians and researchers to gather and potentially disseminate accurate information about contraceptive options.

Table of Contents

Glossary

Contraception-related Terms:

Contraception (a.k.a. Birth Control): the deliberate use of artificial methods or other techniques to prevent pregnancy as a consequence of sexual intercourse.[1]

Copper Intrauterine Device (IUD): a plastic, T-shaped birth control device containing copper coils that can be used for birth control and emergency contraception within five days of unprotected sex. Sold under the brand name 'Paragard', it was approved by the FDA in 1998, and is one of the most effective forms of birth control with a one-year failure rate around 0.8%. The device is placed in the uterus and lasts up to ten years. It may be used by women of all ages regardless of whether or not they have had children. Following removal, fertility quickly returns.[2]

Depo-Provera shot (a.k.a. The Shot): an injectable form of the hormonal medication depot medroxyprogesterone acetate (DMPA), a progestin. It was approved by the FDA in 1992, and is used as a method of birth control. It is also used to treat endometriosis, abnormal uterine bleeding, certain types of cancer, and is used for menopausal hormone therapy. It is typically administered every three months, and provides 98.8% effective contraception for this duration. With typical use, the failure rate per year of use is approximately 5%.[2]

Hormonal Intrauterine Device (IUD): an intrauterine device that releases the hormone levonorgestrel into the uterus. It is used for birth control, heavy menstrual periods, and to prevent excessive growth of the lining of the uterus in those on estrogen replacement therapy. Sold under the brand names Mirena, Liletta, Skyla, and Kyleena, LNG-IUDs are one of the most effective forms of birth control with a one-year failure rate around 0.2%, and have been available in the United States since 2000. The device is placed in the uterus and lasts three to seven years, depending on the brand and dose of progestin. Fertility often returns quickly following removal.[2]

Implanon/Nexplanon implants (a.k.a. Contraceptive Implant) (a.k.a. The implant): a very small plastic rod containing the hormone etonogestrel that is inserted under the skin of a woman's upper arm to provide birth control. It is one of the most effective forms of birth control with a one-year failure rate around 0.05%, and lasts at least three years.[2] One version of contraceptive implant — Implanon — was first used in Indonesia in 1998 and approved for use in the United States in 2006. Nexplanon was developed to eliminate the problem of non-insertion and localization of Implanon by changing the inserter device and making the rod radiopaque. As of January 2012, Implanon is no longer being marketed and Nexplanon is the only available single-rod implant.[3]

IUD (Intrauterine Device): a small, T-shaped birth control device that is inserted into a woman's uterus to prevent pregnancy. IUDs are one form of long-acting reversible birth control (LARC), and can be comprised of copper or levonorgestrel, a progestin.[2]

LARC (Long-acting reversible contraception): methods of birth control that provide effective contraception for an extended period without requiring user action. They include intrauterine devices (IUDs) and subdermal contraceptive implants. They are the most effective reversible methods of contraception because their efficacy is not reliant on patient compliance. Their 'typical use' failure rates, at less than 1% per year, are the same as 'perfect use' failure rates.[2,4]

Oral contraceptives (a.k.a. OCPs) (a.k.a. Birth Control Pills): medications taken by mouth to prevent pregnancy. Two types of female oral contraceptive pill, taken once per day, are widely available: the combined oral contraceptive pill, which contains estrogen and a progestin, and the progestogen-only pill. The first OCPs became available in the United States in the 1960s, and multiple generations of pills have evolved since then. With perfect use, contraceptive pills have a 0.3% failure rate per year of use; with typical use, this rate is approximately 7%. [2]

Pearl Index: the number of contraceptive failures per 100 women-years of exposure. The Pearl Index uses as the denominator the total months or cycles of exposure from the initiation of the product to the end of the study or the discontinuation of the product.[2]

The birth control patch (a.k.a. The patch): a transdermal patch applied to the skin that releases synthetic estrogen and progestin hormones to prevent pregnancy. The first patch available in the United States was the Ortho Evra patch, in 2001. The contraceptive patch has been shown to be as effective as the combined oral contraceptive pill with perfect use (with a 0.3% failure rate per year of use); with typical use, this rate may be as high as 9%.[2]

The vaginal ring (a.k.a. The ring): a polymeric drug delivery device designed to provide controlled release of drugs for intravaginal administration over extended periods of time. The first vaginal ring available in the United States was the NuvaRing, in 2001. The contraceptive vaginal ring is inserted into the vagina and releases etonogestrel (a progestin) and ethinylestradiol (an estrogen), providing contraceptive protection. Vaginal rings come in one size that fits most women. They have been shown to be as effective as the combined oral contraceptive pill with perfect use (with a 0.3% failure rate per year of use); with typical use, this rate may be as high as 9%.[2]

Computation-related Terms:

Amazon SageMaker: a subsidiary of Amazon Web Service, Amazon SageMaker is a fully managed service for developers and data scientists geared toward building, training, and deploying machine learning models. SageMaker provides integrated tools for the entire machine learning workflow in a single toolset, improving efficiency and cost-effectiveness of the machine learning construction process.[5]

API (Application programming interface): an interface or communication protocol between different parts of a computer program intended to simplify the implementation and maintenance of software.[6]

AWS (Amazon Web Service): a subsidiary of Amazon that provides on-demand cloud computing platforms and application programming interfaces (APIs) to individuals, companies, and governments, on a metered pay-as-you-go basis.[7]

GetOldTweets3: a Python program that mines a scroll loader from an internet browser Twitter Search page through calls to a JavaScript Object Notation (JSON) provider, allowing users to conduct manual searches for tweets containing specified keywords or hashtags, from certain usernames, and/or within specific date ranges. The Twitter Search machine harvests all tweets with any combination of the words and/or characters searched for, within the specified time frame, the tweet's permalink, tweeter's username, tweet text, date, number of retweets, number of favorites, number of mentions, and hashtags.[8]

Jupyter Notebook: an open-source web application that allows creation and sharing of documents that contain live code, equations, visualizations and narrative text. Uses include data cleaning and transformation, numerical simulation, statistical modeling, data visualization, and machine learning, among others.[9]

NLP (Natural Language Processing): a subfield of linguistics, computer science, information engineering, and artificial intelligence concerned with the interactions between computers and human languages, in particular how to program computers to process and analyze large amounts of natural language data.[10]

Python: an interpreted, high-level, general-purpose programming language created by Guido van Rossum and first released in 1991 (www.python.org).[11]

R: a free software environment for statistical computing and graphics. R compiles and runs on a wide variety of UNIX platforms, Windows and MacOS. R is widely used at universities and research centers for statistical data analysis, and is now one of the most used software platforms for omics data analysis and its visualization (https://www.r-project.org/).

Twitter: an American microblogging and social networking service where users post and interact with messages known as "tweets". Registered users can post, like, and retweet tweets, but unregistered users can only read them (www.twitter.com).

Acknowledgments

Computing for this project was funded by a grant from the Amazon Web Service (AWS) Cloud Credits for Research Program, to Principal Investigator Allison Merz. AWS had no role in the design or conduct of the study; collection, management, analysis, or interpretation of the data; or preparation, review, or approval of this report.

Tremendous thanks to Paul Avillach, MD, PhD and Deborah Bartz, MD, MPH for their mentorship and support; to Alba Gutierrez Sacristan, PhD for her teaching, guidance and hard work on this project; and to the Harvard medical students Melody Huang, Annie Morgan, Chloe Li, Tomi Ojo, Sofia Sandoval, Kimberly Schaefer, Anabel Starosta Natalie Williams, and Sonya Ye for serving as tweet-reviewers.

Introduction

There are over 74 million women of reproductive age (15–49) in the United States.[12] Approximately 52 million of these (70%) are at risk of unintended pregnancy—that is, they are sexually active with male partners and do not want to become pregnant, but could become pregnant if they and their partners fail to use effective contraception.[13] It is thus no surprise that greater than 99% of women aged 15–44 who have ever had sexual intercourse have used at least one contraceptive method,[14] and over 60% of all women of reproductive age are currently using a contraceptive method.[15] Even accounting for having multiple children, the average woman must use contraceptives for close to four decades of her life.

The consequences of ineffective use and non-use of contraception are important: higher rates of unintended pregnancy are associated with increased rates of abortion,[16] and among continued pregnancies, with less optimal perinatal behaviors, adverse maternal outcomes, premature and low-birth-weight infants, and decreased likelihood of breastfeeding.[17–19] Though rates of unintended pregnancy and abortion in the United States are currently lower than they have been for several decades, even today, nearly half of pregnancies in the United States are unplanned or mis-timed,[16] a statistic that is thought to be due to inconsistent or incorrect use of contraception and an over-reliance on less effective contraceptive methods.[20] Indeed, the decline in unintended pregnancies in recent years has been credited largely to increased contraceptive use, particularly the increasing use of highly effective long-acting reversible contraceptive (LARC) methods such as intrauterine devices (IUDs) and implants, whose *typical use* and *ideal use* Pearl indices (i.e. the percentage of women who will become pregnant over one year of use) are equal, at 0.05-0.8, depending on the specific method[2]. This is compared to short-acting reversible contraceptive (SARC) methods such as the oral contraceptive pill (OCP), the contraceptive patch, the vaginal ring, and the depo shot, which have *ideal use* Pearl indices of less than one, but *typical use* indices of 5-9 because they require daily, weekly, or monthly action by their users (see **Figure 1A)**.[2,16] Despite the greater efficacy, ease-of-use, and high satisfaction rates with LARC methods and relatively worse reliability of SARC methods, the OCP remains the most commonly-used method, and LARC remains under-utilized, though its use is increasing (see **Figure 1B**).[14,20,21] This is thought to be related to lack of knowledge about LARC's risks and benefits and variations in access to it and attitudes toward it, potentially related to age, intrauterine contraception's complex history, and sociocultural norms.[22–24]

Importantly, contraception is unlike other classes of medications in that individuals' method choice is often more strongly influenced by their social network, as compared to other medication decisions that factor in traditional clinical considerations more heavily.[25] Qualitative studies have found that social networks are often the dominant and most trusted source of information about contraception among United States women,[26,27] particularly women under age twenty.[28] Women considering their contraceptive options have been shown to value anecdotal information from friends and family over information from health professionals,[29] with many women specifically utilizing or rejecting methods based on opinions and experiences of members of their social network.[30] Furthermore, women's contraceptive use has been shown to be associated with their perceptions of friends' and family members' contraceptive use and attitudes,[31][32] and among adolescents, having a higher proportion of classmates using contraception has been associated with increased likelihood of an individual using contraception.[33]

Social media now plays a prominent role in the lives and social networks of reproductive-age individuals in the United States and globally, including changing the way many young people communicate and obtain information.[34] 36% of online adults aged 18-29 use Twitter,[35] and because of the prevalence of this network, it offers a unique window into conversations reflecting social norms, behavioral intentions, and sentiments of the population that uses it.[36] Twitter has thus become increasingly recognized as an important source of data for social network analysis, in addition to medical topics ranging from assessing patients' understanding of their conditions and therapies[37,38] to detection of illness and adverse medication effects.[39,40] Furthermore, in recent years, we have seen the profound power that this platform has in our society, from the initiation and spread of the #MeToo movement[41] to its role in setting the agenda for the current US government.[42,43] It is thus no surprise that tweets impact behavior: a significant body of marketing research has demonstrated the influence that "electronic word-of-mouth" and brand mentions on Twitter have on users' behavior.[44–46] In one study, over half of Twitter users reported that they have taken action (searching for, reading about, and purchasing) after seeing brand mentions in Tweets.[47] And in the healthcare realm, a meta-analysis of social networking site-based interventions found such interventions to be effective in promoting health behavior change.[48]

Despite a well-recognized understanding of the social nature of contraceptive decision-making and the prominence of social media in reproductive-age individuals' lives, the extent to which different contraceptive methods are discussed and how they are portrayed on Twitter have not been characterized. Unlike Facebook and Instagram, Twitter is primarily text-based and predictably-sized, with a hard character limit on each post (140 characters until November 2017, and now 280 characters), characteristics that make it well-suited for analysis of unstructured text. Natural Language Processing (NLP) is a type of machine learning in which computers process and analyze large amounts of natural language data. NLP has become a valuable tool for studying the sentiments of healthcare consumers: NLP sentiment analysis in particular has been used to study topics ranging from detection of mental illness to attitudes toward the US healthcare system and human papillomavirus (HPV) vaccination,[49–54] and greater exposure to positive sentiment in online conversations about varenicline has been shown to be associated with a greater likelihood that smokers will choose to use varenicline in a quit attempt.[55]

Importantly, analyzing attitudes toward contraceptive methods with Twitter as a data source may provide improved insight into the attitudes of a population that has traditionally been challenging to engage in research, related to a topic that is often-stigmatized. Indeed, social media has been used to aid in contraceptive counseling[56] and to elicit opinions and attitudes toward family planning among adolescents and young adults, and prior studies have shown that reproductive-age individuals post, seek and are influenced by social media content related to contraception and reproductive health.[57–60] Tweets provide unfiltered sentiment analysis - they may function as a proxy for the mindset of patients considering their contraceptive options, but without the influences or biases that may be present with more traditional research methods such as surveys or in-person interviews, and they are freely available to anyone with internet access. We therefore hypothesize that Twitter will be a valuable data source for analyzing large-scale attitudes toward different contraceptive methods, and that the use of NLP will discern the predominant sentiments associated with different methods since Twitter's founding in March 2006.

The purpose of the current study is thus to explore the attitudes toward different contraceptive methods on Twitter over the past 13 years (March 2006 to December 2019). More specifically, we investigate the portrayal of reversible, prescription contraceptive methods with typical-use Pearl Indices of <10 pregnancies per 100 woman-years, including LARC (the Copper IUD, Levonorgestrel IUDs, and non-specified IUD type; and Implanon/Nexplanon implants) and SARC methods (oral contraceptives; the contraceptive patch; the vaginal ring; and the depo-provera injection) in English language tweets.

Methods

Tweet collection and filtering

We searched for tweets between March 21, 2006 (the date of the first-ever tweet) and December 1, 2019 mentioning reversible, prescription contraceptive methods for which the Pearl index was less than 10%, i.e., copper and levonorgestrel IUDs; the contraceptive implant; oral contraceptive pills; the contraceptive patch; the vaginal ring; and the depo-provera shot (**Figure 1A**).[1] We developed a list of search terms beginning with the names of these contraceptive methods and expanded it to include grammatical and punctuation variations, brand names, and colloquial names for each method, including slang terms encountered on urbanthesaurous.org. The list of search terms was revised iteratively to include only terms that reliably returned tweets related to the desired contraceptive method. The full list of search terms used (including both those that were initially tried and those that were ultimately used) can be found in **Supplementary Table 1.** We searched for tweets about IUDs within three IUD categories (IUD, LNG-IUD, Copper IUD) because many tweets mentioned "IUD" keywords without specifying "copper" or "levonorgestrel/hormonal."

To collect tweets, we used the Python package GetOldTweets3.[8] Once the initial full collection of tweets about contraception was assembled, duplicate tweets, tweets in which the username was a keyword (e.g. @NuvaRingLawyer), and tweets containing the phrases "male contraception," "male contraceptive" "male birth control," "emergency contraception," "emergency contraceptive," and "emergency birth control" were removed because though some tweets harvested with our list of search terms discussed male contraception and emergency contraception (e.g. "when will a male birth control pill come out? It's time…" and "anyone know where I can get a prescription for emergency contraception?"), *a priori* we did not intend to study these topics (**Figure 2**, Steps **B-E**). The number of tweets within each of the eight categories (IUD, LNG-IUD, Copper IUD, Implant, Pill, Patch, Ring and Shot) was quantified at this point of the analysis (**Figure 2,** Step **E**). Note that tweets mentioning more than one method counted toward each mentioned method's total tweet count, meaning that a single tweet could contribute to multiple method classes (for example, the tweet "Have had my IUD for 2 years now and never going back! So much better than the pill" counted toward both "IUD" and "Pill"). Mann Kendall non-parametric tests were used to test for monotonic trends in the number of tweets about each contraceptive method over time.

Automated sentiment analysis

Amazon Comprehend is an NLP service that uses machine learning to find insights and relationships in text using pre-trained models.[61][62] The Sentiment Analysis API interprets unstructured text and returns the overall sentiment of a text (positive, negative, neutral, or mixed) with a confidence score (0-1, 1 being 100% confident) for that sentiment and tweet. To enable automated sentiment analysis of tweets about each birth control method, we narrowed our collection of tweets to include only tweets mentioning a single contraceptive method (Step **F** in **Figure 2**). This was done because automated analysis of tweets mentioning more than one method, such as "the hormonal IUD is amazing!! Way less hormones than the pill and they're localized so much lower chances of side effects" would recognize that this tweet was positive, but would not enable us to discern whether the tweeter felt positively about either or both methods mentioned in the tweet; thus, such tweets were removed by deleting all tweets containing keywords from more than one category (see **Supplementary Table 2**). We used Chi square tests to compare proportions of positive, neutral and negative tweets. To examine the sentiments with respect to each contraceptive method over time, we grouped tweets within each contraception category by year and used Mann Kendall non-parametric tests for monotonic trends in the number of positive, neutral and negative tweets over time.

Manual validation of automated sentiment analysis

In order to test the accuracy of our NLP sentiment analysis on our collection of tweets, we created a "gold standard" collection of tweets based on manual sentiment analysis of tweets by a group of human reviewers (see **Figure 3**, Steps **1-2** and **Supplementary Figure 1**). 1000 tweets were randomly selected from the collection of tweets eligible for sentiment analysis (i.e. all filtered, non-duplicate, single-method tweets). Two groups of five female medical student reviewers each manually analyzed the sentiment of 500 tweets following the guide in **Supplementary Document 1**. Independently and blinded from one another, reviewers determined whether each tweeter felt positively, negatively, neutral, or had mixed emotions about the birth control method mentioned in the tweet. Reviewers also had the option of marking a tweet as a "false positive," meaning that the tweet did not actually mention contraception at all (e.g. "I love going to home depo bc they have free wifi," which was harvested because it contained the search term "depo bc," a term that predominantly returned tweets about the depo shot; another 'false positive' example is the tweet "I think she has Parkinson's disease bc pill rolling tremors are part of that disorder," which was harvested because it contained the search term "bc pill," a term that predominantly returned tweets about the birth control pill). During this manual curation process, we detected two classes of tweets, one related to male contraception and other related to emergency contraception, which were outside of the scope of this project. We thus decided to include an additional filter to exclude these classes of tweets from the entire collection (steps **B** and **C** in **Figure 2**). Because these steps were added after the random extraction, this narrowed the 1000 tweet sample validation to 973 (485 tweets in one group and 488 in the other).

The results of the two groups' manual analysis were consolidated, and if at least three out of five reviewers labeled a tweet as the same sentiment, that result was deemed the "gold standard." The gold standard tweet collection thus included all tweets deemed "positive," "negative," "neutral," or "mixed emotion," with respect to the contraceptive method mentioned in the tweet, in addition to tweets deemed "false positive" (meaning the tweet was not about contraception) by the majority of reviewers (i.e. at least three out of five agreed). Tweets about which a majority of reviewers did not agree on the sentiment were not included in the gold standard. For example, if all five reviewers deemed the tweet "I love my nexplanon!" positive, it would be considered to be positive and be included in the "gold standard" set of tweets. By contrast, if the tweet "birth control pill got me feelin feels" was deemed mixed by two reviewers, neutral by two reviewers, and positive by one reviewer, it was not included in the "gold standard" set of tweets due to lack of agreement on its sentiment.

Interrater reliability (via Fleiss' Kappa score) and observed agreement were calculated to assess agreement among each group of five reviewers. We then compared the NLP-generated categorization to our "gold standard" set of tweets, and derived the sensitivity and specificity of the automated sentiment analysis at different confidence levels (all levels, 80%, 90%, and 95%) for detecting positive, neutral, mixed, and negative sentiments toward the contraceptive method mentioned in a tweet.

Workflow, data sharing and IRB exemption

Our overall study workflow is outlined in **Figure 3**. Tweet extraction was done with Python 3, and tweet filtering and analyses were done with R version 3.6.1. Tweets were collected, grouped, and cleaned in an Amazon Sagemaker notebook instance using Jupyter notebooks. All tweets were stored in an Amazon S3 bucket. The data and analysis code are available at https://github.com/hms-dbmi/contraceptionOnTwitter, and the initial steps of our analysis can be replicated by launching the code here: https://tinyurl.com/cleanTweetsMyBinder.

This study was reviewed and deemed IRB-exempt by the Harvard Medical School Institutional Review Board on the basis of its being non-human-subjects research.

Results

Tweet collection, filtering, and characterization

The initial harvest of tweets containing any search term yielded 989,627 tweets between March 21, 2006 and December 1, 2019 (**Figure 2,** Step **A**). This collection of tweets was narrowed to 838,739 total tweets mentioning at least one contraceptive method after removing tweets mentioning male contraception, emergency contraception, tweets that were collected because the username contained a search term, and duplicate tweets (out of which the earliest tweet was maintained in the collection) (**Figure 2,** Steps **B-E**). The number of tweets removed from each contraceptive class at each filtering step are displayed in **Supplementary Table 3**.

The 838,739 total tweets mentioning at least one contraceptive method were produced by 424,439 users. The median number of tweets mentioning contraception per user was 1 (IQR 1-2), and 123 users tweeted about contraception more than 100 times. The most commonly tweeted-about method was the IUD (47.4%), followed by the shot (15.4%), the pill (11.6%), the implant (11.2%), the ring (7.7%), the copper IUD (2.9%), the patch (2.0%), and lastly, LNG-IUDs (1.9%); LARC methods were mentioned more than SARC methods (58%, vs. 42%) (**Figure 4A**).

The number of tweets about contraception generally increased over time, with the smallest number of tweets in 2007 (35, 0.004 % of all tweets) and the largest in 2019 (187,612, 22.4% of all tweets) (**Table 1**, **Figure 4B**). Out of all tweets collected, the proportion of tweets mentioning LARC methods (all IUDs and the implant) generally increased over time, with the smallest proportion in 2009 (1,460, 22.4%), and the largest proportion in 2019 (143,101, 76.3%). Conversely, the proportion of tweets mentioning SARC methods (pill, patch, ring and shot) was greatest in 2009 (5,055, 77.6%) and smallest in 2019 (44,511, 23.7%) (**Figure 4C**). With respect to individual methods, the number of tweets per year mentioning all LARC methods and the pill trended up over time (all $p< 0.05$), while there was no monotonic upward or downward trend in the number of tweets about the patch, the ring or the shot over time (all $p>0.05$) (**Table 1**).

Manual validation of automated sentiment analysis

There were 889 of the 973 manually reviewed tweets (91.4%) that had agreement of ≥3 reviewers and were included in the gold standard collection, which is publicly available and can be downloaded at https://github.com/hms-dbmi/contraceptionOnTwitter/blob/master/finalGoldStandard.csv. In 403 (45.3%) of the 889, all 5 reviewers agreed on the tweet's classification; 4 reviewers agreed about 229 tweets (25.8%); and 3 agreed about 257 (28.9%). 331 (37.2%) of tweets in the gold standard collection were neutral, 249 (28.0%) were negative, 185 (20.8%) were positive, 95 (10.7%) were mixed, and 29 (3.3%) were false positives. Interrater reliability between the five reviewers in each group was moderate to substantial, with Fleiss' Kappa scores of 0.632 and 0.534 for the two groups. Based on the gold standard 889 tweets' sentiments, the sensitivity and specificity of Amazon Comprehend's sentiment analysis results with 95% confidence were 67% and 92%, 74% and 88%, 83% and 72%, and 0% and 98% for detection of positive, neutral, negative, and mixed sentiments toward the contraceptive method mentioned in a tweet, respectively (**Table 2**).

Automated sentiment analysis

Out of the 838,739 tweets mentioning at least one contraceptive method, a total of 665,064 tweets mentioned a single contraceptive method and were thus eligible for sentiment analysis. To characterize trends in attitudes toward each contraceptive method over time, we looked at tweets interpreted as positive, neutral, or negative by Amazon Comprehend with ≥95% confidence based on our finding that this confidence level led to optimal sensitivity and specificity for the true sentiment toward the contraceptive method mentioned in a tweet (see **Table 2**). Because Amazon Comprehend did not detect mixed emotion tweets reliably (0% sensitive based on our manual gold standard) and because they made up a small proportion of all tweets with sentiment (10.7% of tweets in the gold standard collection), mixed emotion tweets were excluded from our analysis of trends in sentiments over time.

Of the 160,713 tweets with ≥95% confidence score sentiment (24.17% of all tweets subjected to sentiment analysis), the greatest proportion (40.66%) was negative (**Table 3**). With respect to individual methods, the greatest proportion of positive tweets occurred in tweets about the Copper IUD (3,091, 30.37%) while the smallest occurred in tweets about the Patch (335, 7.3%). The greatest proportion of negative tweets occurred in tweets about the Shot (18,805, 61.49%) while the smallest occurred in tweets about hormonal IUDs (590, 15.82%). Overall, there were significantly more positive tweets about LARC methods compared to SARC methods (19.65% vs. 10.21%, $p<0.05$) (**Table 3, Supplementary Tables 5-6**).

In terms of changes and trends over time, there were significant upward trends in the number of both positive and negative tweets about all LARC methods and the pill. There were less clear trends in the number or proportion of positive, neutral or negative tweets about non-LARC methods, though there was a trend toward increasing numbers and proportions of positive tweets about the ring in 2018-2019, and tweets about the shot were predominantly and overall relatively stably negative over time (**Figure 5).**

Discussion

Many factors have been shown to influence the popularity of different contraceptive methods, including product-related factors (for example, availability, marketing, and media coverage of adverse effects and complications), provider-related factors (for example, provider attitudes toward different methods and prevalence of clinicians trained to insert and remove LARCs), and patient-related factors (for example, direct-to-consumer marketing, access, and word of mouth).[24,63] Importantly, the ‘word of mouth’ piece is influenced by all of the others in the list preceding it: for example, FDA approval of a new LARC device, a conversation with a healthcare provider, or a commercial advertising the NuvaRing may all trickle down into consumer discussion, and a significant body of research has shown that such discussion often heavily influences patients’ contraceptive decision-making.[25–33]

This study describes a novel data source in this regard: Twitter offers a window into the minds of its users discussing contraception in a social setting that has become increasingly influential for reproductive-age individuals in the 21st century. We collected English language tweets mentioning reversible, prescription contraceptive methods between March 2006 and December 2019 and found that the number of tweets mentioning contraception each year has increased nearly three hundred-fold since 2007. Furthermore, we found that the proportion of tweets about LARC compared to SARC methods has gradually increased over time since 2009, and in sentiment analysis, there were significantly more positive tweets about LARC methods compared to SARC methods, though the greatest proportion of all tweets about contraceptive methods was negative.

Importantly, we observed patterns in both the overall numbers and in the proportions of positive vs. negative tweets that we believe may reflect the regulation, advertising, availability, use, and satisfaction rates of individual methods, though we did not investigate causal relationships between historical events and tweet volume or content in this analysis. Regarding the overall number of tweets about all contraceptive methods, the number of annual tweets mentioning a contraceptive method gradually increased between 2006 and 2011, peaking in 2012 before leveling off for several years

(**Table 1, Figure 4B**). Though the peak in contraception-related tweets in 2012 is likely multifactorial, we believe that it could be at least in part related to the passage of the "contraception mandate" - the provision of the Affordable Care Act (ACA) requiring private health plans to cover contraceptive methods without copayments, deductibles or other out-of-pocket costs which subsequently took effect for millions of Americans in January 2013[64]. In 2018-2019, the annual number of tweets mentioning all methods of contraception increased substantially once again (**Table 1, Figure 4B**). Shortly before this surge, in October of 2017, the Trump administration expanded exemptions to the contraceptive mandate, restricting coverage of contraception[65]; shortly thereafter, blocks to funding for Title X Family Planning Programs [66] and Planned Parenthood [67] further aimed to restrict access to contraception and family planning services, all contributing to increasing political tension surrounding access to women's reproductive healthcare and the growth of a global movement for women's empowerment and gender equity. This growing modern feminist movement spread widely among younger generations - particularly on social media - in 2018-2019, with 2018 having been called "the year of the woman."[68] Contraception has played a central role in women's empowerment movements throughout history, and it is reasonable to imagine that this surge in online discourse about it is attributable - at least in part - to the current larger movement for gender equity. To this end, the number of tweets about the IUD in particular has surged dramatically over the past several years; that the IUD has become a symbol of the modern feminist movement[69] lends further creed to the notion that the surge in tweets about contreception in recent years may be related to this global phenom. Also interestingly, the number of tweets specifically mentioning the contraceptive pill roughly mirrors the number of tweets about all contraceptive methods, with a peak in 2012, nadir in 2015, and significant upward trend in 2017-2019. This pattern could be reflective of the broader historical context above, considering that the pill is often thought of as the quintessential contraceptive method, with “birth control” often being equated with “the pill” colloquially.

With respect to tweets about specific contraceptive methods, especially LARC, there are additional observations we believe could be related to method availability, advertising, use and satisfaction. The increasing number of tweets and increasing proportions of both positive and negative tweets about LARC methods (**Table 1, Figure 4C, Figure 5**) can likely be attributed to increasing LARC use over the past decade[15]; it logically follows that if more people are using a method, they are more likely to (a) tweet about it, and (b) share their opinions - both good and bad. Several factors underlying increasing LARC use include increasing availability of LARC methods,[13,15] increasing numbers of providers who are trained to insert LARC methods,[24,70–72] and increases in direct-to-consumer advertising of LARC methods, with LARC surpassing the OCP to become the most heavily-promoted class of contraception in 2012.[73] Perhaps related to these phenomena, the number of annual tweets mentioning OCPs peaked in 2012 and then gradually declined for several years following, while 2013 saw a spike in tweets mentioning LNG IUDs and a steady increase in mentions of all LARC methods in subsequent years (**Table 1, Figure 4B-C)**. Looking at hormonal IUDs specifically, the increasing number and proportion of *positive* tweets in recent years (the LNG-IUD and the vaginal ring were the only methods for which the number of positive tweets ever surpassed both neutral and negative; see

Figure 5) could be related to their increasing availability, use, and likability: the lower-cost 52mg levonorgestrel IUD (Liletta) was approved in 2015, and drove the cost of competitor LNG-IUDs down, increasing accessibility of these methods.[74] Concomitantly, hormonal IUDs have the highest satisfaction rates (70% “very satisfied”) and lowest one-year discontinuation rates (12.5%) of all reversible contraceptive methods.[75]

Patterns in tweets about the ring, patch and shot (**Figure 5**) may also reflect historical events and consumers’ experiences. In 2011, the FDA issued a black box warning for thrombotic risk associated with the Ortho Evra patch[76], and in this same year, both the overall number and the number of negative tweets about the patch peaked. In an opposite trend, in 2018, the FDA approved a new 1-year contraceptive ring - Annovera[77] - and in that same year, the number of positive tweets mentioning contraceptive rings increased substantially, subsequently surpassing the numbers of neutral and negative tweets in 2019. Finally, the stably negative majority of tweets about the shot can be understood in the context of well-characterized dissatisfaction with the depo-provera injection: with a 43% one-year discontinuation rate, this method is consistently associated with the most side effects and the lowest satisfaction of all reversible contraceptive methods.[75,78]

In addition to the major findings with respect to individual methods described above, the high prevalence of negative tweets about all contraceptive methods is striking. This may be due in part to a negativity bias related to individuals’ experiences with contraception and contemplation of their contraceptive options. Negativity bias is the notion that, even when of equal intensity, things of a more negative nature (for example, unpleasant thoughts or emotions, or harmful/traumatic events - perhaps cramping, heavy bleeding, a painful procedure, or the prospect of an unintended pregnancy, in this case) have a greater effect on one's psychological state and processes than neutral or positive things.[79] Across an array of psychological situations and tasks, adults display a propensity to attend to, learn from, and use negative information far more than positive information[80] and neurophysiologically, event-related potential and functional magnetic resonance imaging studies have shown stronger neural responses to negative stimuli compared to positive ones.[81–83] Our finding that negative tweets such as “I hate the stupid depo shot, I’ve gained 20 lbs in 6 months and I cry all the time now” are more common than tweets such as “I’ve been happy with the depo shot so far!” is in keeping with our human tendency to be aware of and share more negative experiences than positive ones. Furthermore, negativity bias has been shown to be particularly relevant in decision-making: when presented with the prospect of either gaining or losing something, potential costs are often more heavily considered than potential gains; in other words, people fear the consequences of the negative outcome more than they desire the potential positive outcome, even when the two possibilities are equivalent.[84] In the case of womens’ contraceptive experiences and decision-making as portrayed on Twitter, this bias is manifested insofar as tweets such as “Terrified about my nexplanon appointment tomorrow. May or may not go.” and “I want an IUD but I heard it hurts a lot...so no” were far more common than tweets such as “Sure, taking a pill every day is annoying, but I’m glad I don’t have to worry about pregnancy!” Related to our human negativity bias more broadly, it is worth noting the possibility that Twitter as a platform highlights negative attitudes more than positive ones, not just about contraception, but in tweets more broadly; though this has not been explicitly investigated with tweets, there is evidence that negative valence is associated with increasing virality of online content.[85] Finally, it is also worth noting that based on our manual gold standard collection of tweets, the sensitivity of our NLP algorithm for negative sentiment toward a method was 83%, compared to only 67% for positive sentiment toward a method. It is thus possible that our NLP analysis over-estimated the proportion of negative tweets, although among the 889 tweets in the gold standard collection, a greater proportion of tweets were negative compared to positive (28% vs. 20.8%).

Finally, an additional prominent trend in our sentiment analysis was a regression toward greater proportions of tweets with emotional valence over time (i.e. more positive and negative, fewer neutral) among nearly all contraceptive methods. Several hypotheses could explain this observation: first, in our increasingly polarized world, social media sites are becoming recognized as “echo chambers” in which individuals post more and more polarized content generally, ranging from the political to the personal.[86] This ‘echo chamber’ phenomenon may be related to the idea of emotional contagion, or the notion that one person's (or post’s) emotions directly trigger similar emotions and behaviors in others; indeed, prior experimental research using Facebook has shown that when users are exposed to more negative posts, their own posts become more negative, and vice versa.[87] Recognizing that either or both of these phenomena is potentially at play, the clear trend toward more tweets with emotional valence about all contraceptive methods becomes more understandable.

In considering the potential impact of this work, it is important to acknowledge its limitations. First, we by no means collected *all* tweets that mention the contraceptive methods we studied: we were unable to harvest tweets with certain brand names of medications due to their also being names, thus returning many tweets unrelated to contraception (e.g. “Yasmin”), or tweets with mis-spelled terms (e.g. birht control pill). Second, a small percentage of tweets were not actually about contraception: extrapolating from the 3.3% of the 973 tweets that were manually reviewed that were false positives, we estimate that approximately 27,000 tweets out of our greater-than-838,000-tweet collection are not actually about contraception. Third, we were not able to include tweets’ images, gifs or linked web pages in our sentiment analysis. Fourth, our data may be subject to bias in that Twitter users are predominantly democratic, and a minority of users account for the majority of tweets.[88] Regarding the former point, we do not have demographic or political information for our users, as this information is not typically posted on users’ profiles, making interpretation of the potential generalizability of our results challenging. Regarding the latter, however, we do know that the median number of tweets per user in our sample was 1 (IQR 1-2), dampening any potential concern that a small number of users could be heavily biasing our sample. Finally, our NLP sentiment analysis was limited: to look at sentiments toward specific methods, we excluded all tweets mentioning more than one method; only 24% of tweets had NLP-assigned sentiments with ≥95% confidence (though this was still more than 160,000 tweets, this sample is significantly smaller than our complete sample of tweets); and our sentiment analysis program had a sensitivity and specificity of only 67% and 92%, 74% and 88%, and 83% and 72% for detection of positive, neutral, and negative sentiments toward the contraceptive method mentioned in a tweet, respectively. Factors underlying these less-than-perfectly accurate categorization methods include the program's inability to detect sarcasm (e.g. "just got my birth control shot directly on my sunburn.. Lovely!" was interpreted as positive, though we can tell the true sentiment of this tweet is probably negative), in addition to its inability to distinguish between the attitude toward the method mentioned in the tweet and the tweet's overall sentiment (e.g. "I'm so excited to get my nexplanon out and reset my body" was interpreted as positive, though we can tell that the sentiment toward the nexplanon is negative). As discussed above, it is possible that our NLP analysis over-estimated the proportion of negative tweets based on its greater sensitivity for negative sentiment toward a contraceptive method, though the gold standard collection had a greater proportion of negative tweets compared to positive (28% vs. 20.8%).

Despite these limitations, this study has several notable strengths. First and foremost, to our knowledge, it is the first large-scale study of the portrayal of reversible, prescription contraceptive methods on any form of social media, a potentially significant influential factor in contraceptive decision-making for reproductive-age individuals. Second, we validated the performance of Amazon Comprehend's NLP sentiment analysis algorithm for our data based on a gold standard manual sentiment analysis by ten human reviewers, among whom agreement was very good (91.4% of tweets had majority agreement). Third, its broad scope enables the investigation of not just a single contraceptive method or class of methods, but *all* of the reversible, prescription methods. Fourth, traditional research methods investigating attitudes toward treatment options such as focus groups and face-to-face interviewing are time consuming, vulnerable to bias depending on the fellow group members and/or interviewers, and costly. By contrast, tweets are free from potential influence of interviewers or group members and are readily available for free. To this end, we have made our database of tweets publicly available, setting up future investigators to build upon our work. Broadly speaking, mining and interpreting tweets is a way for researchers to be front-facing with contraceptive users - key stakeholders in contraceptive development and satisfaction. This could be an important way for researchers and clinicians to better understand the needs, desires, and frustrations of contraceptive methods, potentially guiding future research directions and programs.

Along these lines, this work sets the stage for several potential areas of future investigation. First, analyzing the geographic distribution of tweets in our sample would be enlightening: though tweets' locations were not available, the location information of approximately 60% of the twitter users in our sample can be collected from their profiles, and we are in the process of collecting this data. Second, the ideal goal of this research would be to discern whether method-mentions and sentiments correlate with or influence real-world method use by the tweeters. To this end, we could, after collecting the geographic data described above, correlate tweets produced by users from specific geographic regions with method use in those regions based on insurance claims data; or, for that matter, one could tweet at the tweeters directly and ask about their contraceptive use. Third, a more advanced NLP model could better characterize and interpret the tweets about contraception. Feasible analysis goals could include distinguishing between provider- vs. consumer-produced tweets, recognizing attitudes toward different contraceptive methods with greater nuance and accuracy than simply “positive, negative or neutral”, interpreting tweets mentioning more than one method, and detecting sarcasm, as prior work has done.[89] Fourth, one of the prominent observations in this study was the regression toward increasing emotional valence (i.e. both positive and negative) in all tweets about contraception over time. It is unclear whether this trend reflects polarization among Twitter users discussing contraception, or if Twitter as a whole has become more polarized. An additional route of investigation could include performing sentiment analysis on a random sample of tweets about an unrelated topic to determine whether similar trends are seen in tweets unrelated to contraception. Finally, beyond these information-gathering routes, this work sets the stage for potential interventions: knowing that consumers discuss contraception on twitter, can we use this information to help individuals discover methods that better suit their needs? Indeed, prior research has demonstrated that a facebook-based guide about contraception as an adjunct to in-office counseling increased patient preference for LARC.[56] Smartphones are now able to detect depressive symptoms in their users and suggest mental health hotlines[90]; what if your smartphone detected that you were dissatisfied with your contraceptive method and asked if you’d heard of specific alternative options?

In conclusion, our findings suggest that tweets about contraception are a previously-unexplored, easily accessible, and valuable source of data reflecting attitudes toward contraceptive methods and how they have evolved over the past 13 years. Bearing in mind the limitations of this work, its implications are far-reaching. Particularly considering the potential links between observed trends in tweet volume and sentiment and their historical context, our findings suggest that using Twitter as a data source may be (1) a valid means of assessing individuals' experiences of different contraceptive methods, and (2) a valuable tool for dissemination of accurate information about individuals' contraceptive options with the potential to help people find methods that will best suit their needs and priorities. Our methods and their application to other consumer-generated media have the potential to vastly improve our insight into the attitudes of a population that has traditionally been challenging to engage in research, related to a topic that is often-stigmatized. If indeed we feel that messaging surrounding contraception is important as a way to minimize unintended pregnancies, what can we do at the level of twitter, and social media more broadly, to potentially affect public opinion so that individuals' contraceptive needs are better met? Recognizing that patients have more negative than positive things to say about their contraception, could we more effectively message positive aspects of our various contraceptive methods? Could mining and interpreting tweets about contraception help us identify patients who need improved reproductive healthcare provision? Many people have created tremendously powerful movements via Twitter - from #MeToo to Donald Trump. Could a similar movement about contraception impact individuals' attitudes? This work represents a first step toward a better understanding of the role of social media in contraceptive decision-making, potentially setting the stage for an entirely new realm of investigation and intervention in family planning.

Summary

We used Natural Language Processing to explore attitudes toward different contraceptive methods on Twitter since 2006. We collected 989,627 English language tweets mentioning reversible, prescription contraceptive methods with typical-use Pearl Index of <10 pregnancies per 100 woman-years, including prescription LARC (IUDs and the contraceptive implant) and SARC (oral contraceptive pills; the contraceptive patch; the vaginal ring; and the depo-provera shot) between March 2006 and December 2019. The most commonly tweeted-about method was the IUD (45.9%), and LARC methods were mentioned more than SARC methods (58% vs. 42%). In NLP sentiment analysis validated by a manual sentiment analysis involving ten human reviewers, there were nearly twice as many positive tweets about LARC methods compared to SARC methods (19.65% vs. 10.21%, $p<0.05$), though the majority of tweets was negative (40.66%). We observed potential links between trends in tweet volume and sentiment and their historical context, including their regulation, advertising, availability, and satisfaction. Our findings suggest that using Twitter as a data source may be (1) a new means of assessing individuals' experiences of different contraceptive methods, and (2) a valuable tool for dissemination of accurate information about individuals' contraceptive options with the potential to help people find methods that will best suit their needs and priorities.

References

[1] Centers for Disease Control and Prevention. Contraception | Reproductive Health | CDC [Internet]. 2019 [cited 2019 Sep 1]. Available from: ttps://www.cdc.gov/reproductivehealth/contraception/index.htm.

[2] Hatcher RA. Contraceptive Technology [Internet]. Ardent Media; 2007. Available from: https://play.google.com/store/books/details?id=txh0LpjjhkoC.

[3] Mansour D. Nexplanon(®): what Implanon(®) did next. J. Fam. Plann. Reprod. Health Care [Internet]. 2010;36:187–189. Available from: http://dx.doi.org/10.1783/147118910793048629.

[4] Stoddard A, McNicholas C, Peipert JF. Efficacy and safety of long-acting reversible contraception. Drugs [Internet]. 2011;71:969–980. Available from: http://dx.doi.org/10.2165/11591290-000000000-00000.

[5] Amazon Web Service. Amazon SageMaker [Internet]. Amazon Web Service. 2019 [cited 2019 Oct 20]. Available from: https://aws.amazon.com/sagemaker/.

[6] The Editors of Encyclopaedia Britannica. API COMPUTER PROGRAMMING [Internet]. Encyclopedia Brittanica. 2020 [cited 2020 Jan 21]. Available from: https://www.britannica.com/technology/API.

[7] Amazon Web Service, Inc. Amazon Web Service [Internet]. Amazon Web Service. 2019 [cited 2019 Mar 20]. Available from: https://aws.amazon.com/.

[8] Jefferson Henrique And. GetOldTweets3: A Python 3 library and a corresponding command line utility for accessing old tweets [Internet]. 2019 [cited 2019 Dec 1]. Available from: https://pypi.org/project/GetOldTweets3/.

[9] Jupyter. Jupyter [Internet]. Jupyter. 2019 [cited 2019 Oct 20]. Available from: https://jupyter.org/.

[10] Natural language processing [Internet]. Wikipedia: the free encyclopedia. 2020 [cited 2020 Jan 21]. Available from: https://en.wikipedia.org/wiki/Natural_language_processing.

[11] Python. Python [Internet]. Python. 2020 [cited 2020 Jan 21]. Available from: www.python.org.

[12] United States Census Bureau. Age and Sex Composition in the United States: 2018 [Internet]. United States Census Bureau. 2019 [cited 2020 Jan 25]. Available from: https://www.census.gov/data/tables/2018/demo/age-and-sex/2018-age-sex-composition.html.

[13] Daniels K, Daugherty J, Jones J, et al. Current Contraceptive Use and Variation by Selected Characteristics Among Women Aged 15-44: United States, 2011-2013. Natl. Health Stat. Report. [Internet]. 2015;1–14. Available from: https://www.ncbi.nlm.nih.gov/pubmed/26556545.

[14] Daniels K, Mosher WD. Contraceptive methods women have ever used: United States,

1982-2010. Natl. Health Stat. Report. [Internet]. 2013;1–15. Available from: https://www.ncbi.nlm.nih.gov/pubmed/24988816.

[15] Kavanaugh ML, Jerman J. Contraceptive method use in the United States: trends and characteristics between 2008, 2012 and 2014. Contraception [Internet]. 2018;97:14–21. Available from: http://dx.doi.org/10.1016/j.contraception.2017.10.003.

[16] Finer LB, Zolna MR. Declines in Unintended Pregnancy in the United States, 2008–2011. N. Engl. J. Med. [Internet]. 2016;374:843–852. Available from: https://doi.org/10.1056/NEJMsa1506575.

[17] Cheng D, Schwarz EB, Douglas E, et al. Unintended pregnancy and associated maternal preconception, prenatal and postpartum behaviors. Contraception [Internet]. 2009;79:194–198. Available from: http://dx.doi.org/10.1016/j.contraception.2008.09.009.

[18] Pryor J, Patrick SW, Sundermann AC, et al. Pregnancy Intention and Maternal Alcohol Consumption. Obstet. Gynecol. [Internet]. 2017;129:727–733. Available from: http://dx.doi.org/10.1097/AOG.0000000000001933.

[19] Mohllajee AP, Curtis KM, Morrow B, et al. Pregnancy Intention and Its Relationship to Birth and Maternal Outcomes [Internet]. Obstetrics & Gynecology. 2007. p. 678–686. Available from: http://dx.doi.org/10.1097/01.aog.0000255666.78427.c5.

[20] Guttmacher Institute. Contraceptive Use in the United States: Fact Sheet [Internet]. Guttmacher Institute. 2018 [cited 2019 Jul 22]. Available from: https://www.guttmacher.org/fact-sheet/contraceptive-use-united-states.

[21] Kimberly Daniels, Ph.D., and Joyce C. Abma. Current Contraceptive Status Among Women Aged 15–49: United States, 2015–2017: NCHS Data Brief No. 327, December 2018 [Internet]. Centers for Disease Control and Prevention. 2018 [cited 2019 Nov 1]. Available from: https://www.cdc.gov/nchs/products/databriefs/db327.htm.

[22] Bachorik A, Friedman J, Fox A, et al. Adolescent and Young Adult Women's Knowledge of and Attitudes Toward Etonogestrel Implants. J. Pediatr. Adolesc. Gynecol. [Internet]. 2015;28:229–233. Available from: http://dx.doi.org/10.1016/j.jpag.2014.08.002.

[23] Mestad R, Secura G, Allsworth JE, et al. Acceptance of long-acting reversible contraceptive methods by adolescent participants in the Contraceptive CHOICE Project. Contraception [Internet]. 2011;84:493–498. Available from: http://dx.doi.org/10.1016/j.contraception.2011.03.001.

[24] Hubacher D, Finer LB, Espey E. Renewed interest in intrauterine contraception in the United States: evidence and explanation. Contraception [Internet]. 2011;83:291–294. Available from: http://dx.doi.org/10.1016/j.contraception.2010.09.004.

[25] Dehlendorf C, Levy K, Kelley A, et al. Women's preferences for contraceptive counseling and decision making. Contraception [Internet]. 2013;88:250–256. Available from: http://dx.doi.org/10.1016/j.contraception.2012.10.012.

[26] Blackstock OJ, Mba-Jonas A. Family planning knowledge: The role of social networks and primary care providers as information sources for African American women. American Journal of [Internet]. 2010; Available from:

https://www.tandfonline.com/doi/abs/10.1080/10627197.2010.491060.

[27] Carter MW, Bergdall AR, Henry-Moss D, et al. A qualitative study of contraceptive understanding among young adults [Internet]. Contraception. 2012. p. 543–550. Available from: http://dx.doi.org/10.1016/j.contraception.2012.02.017.

[28] Levy K, Minnis AM, Lahiff M, et al. Bringing patients' social context into the examination room: an investigation of the discussion of social influence during contraceptive counseling. Womens. Health Issues [Internet]. 2015;25:13–21. Available from: http://dx.doi.org/10.1016/j.whi.2014.10.001.

[29] Gilliam ML, Warden M, Goldstein C, et al. Concerns about contraceptive side effects among young Latinas: A focus-group approach [Internet]. Contraception. 2004. p. 299–305. Available from: http://dx.doi.org/10.1016/j.contraception.2004.04.013.

[30] Yee L, Simon M. The Role of the Social Network in Contraceptive Decision-making Among Young, African American and Latina Women [Internet]. Journal of Adolescent Health. 2010. p. 374–380. Available from: http://dx.doi.org/10.1016/j.jadohealth.2010.03.014.

[31] Gayen K, Raeside R. Social networks and contraception practice of women in rural Bangladesh. Soc. Sci. Med. [Internet]. 2010;71:1584–1592. Available from: http://dx.doi.org/10.1016/j.socscimed.2010.08.002.

[32] Valente TW, Watkins SC, Jato MN, et al. Social network associations with contraceptive use among Cameroonian women in voluntary associations. Soc. Sci. Med. [Internet]. 1997;45:677–687. Available from: http://dx.doi.org/10.1016/s0277-9536(96)00385-1.

[33] Ali MM, Amialchuk A, Dwyer DS. Social network effects in contraceptive behavior among adolescents. J. Dev. Behav. Pediatr. [Internet]. 2011;32:563–571. Available from: http://dx.doi.org/10.1097/DBP.0b013e318231cf03.

[34] Miller D, Costa E, Haynes N, et al. How the World Changed Social Media [Internet]. UCL Press; 2016. Available from: https://play.google.com/store/books/details?id=Iyk8DwAAQBAJ.

[35] Greenwood S, Perrin A, Duggan M. Social media update 2016. Pew Research Center [Internet]. 2016;11. Available from: http://downtowndubuque.org/wp-content/uploads/2017/01/Social-Media-Update-2016.pdf.

[36] Asur S, Huberman BA. Predicting the Future with Social Media [Internet]. 2010 IEEE/WIC/ACM International Conference on Web Intelligence and Intelligent Agent Technology. 2010. Available from: http://dx.doi.org/10.1109/wi-iat.2010.63.

[37] Martinez B, Dailey F, Almario CV, et al. Patient Understanding of the Risks and Benefits of Biologic Therapies in Inflammatory Bowel Disease: Insights from a Large-scale Analysis of Social Media Platforms. Inflamm. Bowel Dis. [Internet]. 2017;23:1057–1064. Available from: http://dx.doi.org/10.1097/MIB.0000000000001110.

[38] Pemmaraju N, Utengen A, Gupta V, et al. Social media and myeloproliferative neoplasms (MPN): analysis of advanced metrics from the first year of a new Twitter community:# MPNSM. Curr. Hematol. Malig. Rep. [Internet]. 2016;11:456–461. Available from: https://idp.springer.com/authorize/casa?redirect_uri=https://link.springer.com/article/10.100

7/s11899-016-0341-2&casa_token=XzIyY_yz7s8AAAAA:v1c6vCGXOdN4Qp89UTgWK4L1ZLCEDcx7PIwoQmliqBT2FtBoHznmrAXbeh9J3WBuz-YeBVmSAW8uLoxV.

[39] Ru B, Harris K, Yao L. A Content Analysis of Patient-Reported Medication Outcomes on Social Media. 2015 IEEE International Conference on Data Mining Workshop (ICDMW) [Internet]. ieeexplore.ieee.org; 2015. p. 472–479. Available from: http://dx.doi.org/10.1109/ICDMW.2015.150.

[40] Coppersmith G, Dredze M, Harman C. Quantifying mental health signals in Twitter. Proceedings of the workshop on computational linguistics and clinical psychology: From linguistic signal to clinical reality [Internet]. aclweb.org; 2014. p. 51–60. Available from: https://www.aclweb.org/anthology/W14-3207.

[41] Chicago Tribune. #MeToo: A timeline of events. The Chicago Tribune. 2020 Jan 6;

[42] Williams CB. Social Media, Political Marketing and the 2016 U.S. Election [Internet]. 2018. Available from: http://dx.doi.org/10.4324/9781351105521.

[43] Dunning E. # TrumpStyle: The Political Frames and Twitter Attacks of Donald Trump. The Journal of Social Media in Society [Internet]. 2018;7:205–231. Available from: http://www.thejsms.org/tsmri/index.php/TSMRI/article/view/413.

[44] Jansen BJ, Zhang M, Sobel K, et al. Twitter power: Tweets as electronic word of mouth. J. Am. Soc. Inf. Sci. [Internet]. 2009;60:2169–2188. Available from: http://doi.wiley.com/10.1002/asi.21149.

[45] Kozinets RV, De Valck K, Wojnicki AC, et al. Networked Narratives: Understanding Word-of-Mouth Marketing in Online Communities. J. Mark. [Internet]. 2010;74:71–89. Available from: https://doi.org/10.1509/jm.74.2.71.

[46] Forbes LP. Does Social Media Influence Consumer Buying Behavior? An Investigation Of Recommendations And Purchases. 1 [Internet]. 2013 [cited 2020 Feb 9];11:107–112. Available from: https://www.clutejournals.com/index.php/JBER/article/view/7623.

[47] @amidha. Discovering the Value of Earned Audiences — How Twitter Expressions Activate Consumers [Internet]. Twitter Blog. 2014 [cited 2020 Jan 21]. Available from: https://blog.twitter.com/en_us/a/2014/study-exposure-to-brand-tweets-drives-consumers-to-take-action-both-on-and-off-twitter.html.

[48] Laranjo L, Arguel A, Neves AL, et al. The influence of social networking sites on health behavior change: a systematic review and meta-analysis. J. Am. Med. Inform. Assoc. [Internet]. 2015;22:243–256. Available from: http://dx.doi.org/10.1136/amiajnl-2014-002841.

[49] De Choudhury M, Gamon M, Counts S, et al. Predicting depression via social media. Seventh international AAAI conference on weblogs and social media [Internet]. 2013. Available from: https://www.aaai.org/ocs/index.php/ICWSM/ICWSM13/paper/viewPaper/6124.

[50] Arseniev-Koehler A, Lee H, McCormick T, et al. #Proana: Pro-Eating Disorder Socialization on Twitter [Internet]. Journal of Adolescent Health. 2016. p. 659–664. Available from: http://dx.doi.org/10.1016/j.jadohealth.2016.02.012.

[51] Birnbaum ML, Ernala SK, Rizvi AF, et al. A Collaborative Approach to Identifying Social Media Markers of Schizophrenia by Employing Machine Learning and Clinical Appraisals. J. Med. Internet Res. [Internet]. 2017;19:e289. Available from: http://dx.doi.org/10.2196/jmir.7956.

[52] Zhang H, Wheldon C, Dunn AG, et al. Mining Twitter to assess the determinants of health behavior toward human papillomavirus vaccination in the United States. J. Am. Med. Inform. Assoc. [Internet]. 2020;27:225–235. Available from: http://dx.doi.org/10.1093/jamia/ocz191.

[53] Hawkins JB, Brownstein JS, Tuli G, et al. Measuring patient-perceived quality of care in US hospitals using Twitter. BMJ Qual. Saf. [Internet]. 2016;25:404–413. Available from: http://dx.doi.org/10.1136/bmjqs-2015-004309.

[54] Leis A, Ronzano F, Mayer MA, et al. Detecting Signs of Depression in Tweets in Spanish: Behavioral and Linguistic Analysis [Internet]. Journal of Medical Internet Research. 2019. p. e14199. Available from: http://dx.doi.org/10.2196/14199.

[55] Cobb NK, Mays D, Graham AL. Sentiment analysis to determine the impact of online messages on smokers' choices to use varenicline. J. Natl. Cancer Inst. Monogr. [Internet]. 2013;2013:224–230. Available from: http://dx.doi.org/10.1093/jncimonographs/lgt020.

[56] Kofinas JD, Varrey A, Sapra KJ, et al. Adjunctive social media for more effective contraceptive counseling: a randomized controlled trial. Obstet. Gynecol. [Internet]. 2014;123:763–770. Available from: http://dx.doi.org/10.1097/AOG.0000000000000172.

[57] Altshuler AL, Gerns Storey HL, Prager SW. Exploring abortion attitudes of US adolescents and young adults using social media. Contraception [Internet]. 2015;91:226–233. Available from: http://dx.doi.org/10.1016/j.contraception.2014.11.009.

[58] Mercier RJ, Senter K, Webster R, et al. Instagram Users' Experiences of Miscarriage. Obstet. Gynecol. [Internet]. 2020;135:166–173. Available from: http://dx.doi.org/10.1097/AOG.0000000000003621.

[59] Nguyen BT, Allen AJ. Social media and the intrauterine device: a YouTube content analysis. BMJ Sex Reprod Health [Internet]. 2018;44:28–32. Available from: http://dx.doi.org/10.1136/bmjsrh-2017-101799.

[60] Stevens R, Gilliard-Matthews S, Dunaev J, et al. Social Media Use and Sexual Risk Reduction Behavior Among Minority Youth [Internet]. Nursing Research. 2017. p. 368–377. Available from: http://dx.doi.org/10.1097/nnr.0000000000000237.

[61] Amazon Web Services, Inc. Amazon Comprehend [Internet]. Amazon Web Service. 2019 [cited 2019 Nov 1]. Available from: https://aws.amazon.com/comprehend/.

[62] Sahu P. Analyze content with Amazon Comprehend and Amazon SageMaker notebooks [Internet]. AWS Machine Learning Blog. 2019 [cited 2019 Nov 2]. Available from: https://aws.amazon.com/blogs/machine-learning/analyze-content-with-amazon-comprehend-and-amazon-sagemaker-notebooks/.

[63] Westhoff C, Marks F, Rosenfield A. Residency training in contraception, sterilization, and abortion. Obstet. Gynecol. [Internet]. 1993;81:311–314. Available from:

https://www.ncbi.nlm.nih.gov/pubmed/8423971.

[64] Finer LB, Sonfield A, Jones RK. Changes in out-of-pocket payments for contraception by privately insured women during implementation of the federal contraceptive coverage requirement. Contraception [Internet]. 2014;89:97–102. Available from: http://dx.doi.org/10.1016/j.contraception.2013.11.015.

[65] Laurie Sobel Follow @laurie_sobel on Twitter, Alina Salganicoff Follow @a_salganicoff on Twitter, and Caroline Rosenzweig. New Regulations Broadening Employer Exemptions to Contraceptive Coverage: Impact on Women [Internet]. Kaiser Family Foundation. 2020 [cited 2018 Nov 19]. Available from: https://www.kff.org/health-reform/issue-brief/new-regulations-broadening-employer-exemptions-to-contraceptive-coverage-impact-on-women/.

[66] Cha AE. Is it a gag rule after all? A closer look at changes to Title X funding regarding abortion [Internet]. Washington Post. 2018 [cited 2020 Feb 4]. Available from: https://www.washingtonpost.com/news/to-your-health/wp/2018/05/23/is-it-a-gag-rule-what-changes-to-family-planning-funds-and-abortion-referrals-might-mean/.

[67] Belluck P. Trump Administration Blocks Funds for Planned Parenthood and Others Over Abortion Referrals [Internet]. New York Times. 2019 [cited 4 Feburary 2020]. Available from: https://www.nytimes.com/2019/02/22/health/trump-defunds-planned-parenthood.html.

[68] Schnall M. will be the year of women. CNN. 2018.

[69] Copeland L. From Medical Pariah to Feminist Icon: The Story of the IUD [Internet]. Smithsonian Magazine. 2017 [cited 2020 Feb 1]. Available from: https://www.smithsonianmag.com/science-nature/medical-pariah-feminist-icon-story-iud-180963699/.

[70] Tyler CP, Whiteman MK, Zapata LB, et al. Health care provider attitudes and practices related to intrauterine devices for nulliparous women. Obstet. Gynecol. [Internet]. 2012;119:762–771. Available from: http://dx.doi.org/10.1097/AOG.0b013e31824aca39.

[71] Rubin SE, Fletcher J, Stein T, et al. Determinants of intrauterine contraception provision among US family physicians: a national survey of knowledge, attitudes and practice. Contraception [Internet]. 2011;83:472–478. Available from: http://dx.doi.org/10.1016/j.contraception.2010.10.003.

[72] Ricketts S, Klingler G, Schwalberg R. Game change in Colorado: Widespread use of long-acting reversible contraceptives and rapid decline in births among young, low-income women. Perspect. Sex. Reprod. Health [Internet]. 2014;46:125–132. Available from: https://onlinelibrary.wiley.com/doi/abs/10.1363/46e1714.

[73] Wu MH, Bartz D, Avorn J, et al. Trends in direct-to-consumer advertising of prescription contraceptives. Contraception [Internet]. 2016;93:398–405. Available from: http://dx.doi.org/10.1016/j.contraception.2016.01.010.

[74] Roth LP, Sanders JN, Simmons RG, et al. Changes in uptake and cost of long-acting reversible contraceptive devices following the introduction of a new low-cost levonorgestrel IUD in Utah's Title X clinics: a retrospective review [Internet]. Contraception. 2018. p. 63–68. Available from: http://dx.doi.org/10.1016/j.contraception.2018.03.029.

[75] Peipert JF, Zhao Q, Allsworth JE, et al. Continuation and satisfaction of reversible contraception. Obstet. Gynecol. [Internet]. 2011;117:1105–1113. Available from: http://dx.doi.org/10.1097/AOG.0b013e31821188ad.

[76] Raymond EG, Burke AE, Espey E. Combined hormonal contraceptives and venous thromboembolism: putting the risks into perspective. Obstet. Gynecol. [Internet]. 2012;119:1039–1044. Available from: http://dx.doi.org/10.1097/AOG.0b013e31825194ca.

[77] Nelson AL. Comprehensive overview of the recently FDA-approved contraceptive vaginal ring releasing segesterone acetate and ethinylestradiol: A new year-long, patient controlled, reversible birth control method [Internet]. Expert Review of Clinical Pharmacology. 2019. p. 953–963. Available from: http://dx.doi.org/10.1080/17512433.2019.1669448.

[78] Dianat S, Fox E, Ahrens KA, et al. Side Effects and Health Benefits of Depot Medroxyprogesterone Acetate: A Systematic Review. Obstet. Gynecol. [Internet]. 2019;133:332–341. Available from: http://dx.doi.org/10.1097/AOG.0000000000003089.

[79] Rozin P, Royzman EB. Negativity Bias, Negativity Dominance, and Contagion. Pers. Soc. Psychol. Rev. [Internet]. 2001;5:296–320. Available from: https://doi.org/10.1207/S15327957PSPR0504_2.

[80] Vaish A, Grossmann T, Woodward A. Not all emotions are created equal: the negativity bias in social-emotional development. Psychol. Bull. [Internet]. 2008;134:383–403. Available from: http://dx.doi.org/10.1037/0033-2909.134.3.383.

[81] Zhang H, Qu C. Emotional, especially negative microblogs are more popular on the web: evidence from an fMRI study. Brain Imaging Behav. [Internet]. 2018; Available from: http://dx.doi.org/10.1007/s11682-018-9998-6.

[82] Kark SM, Slotnick SD, Kensinger EA. Forgotten but not gone: FMRI evidence of implicit memory for negative stimuli 24 hours after the initial study episode. Neuropsychologia [Internet]. 2020;136:107277. Available from: http://dx.doi.org/10.1016/j.neuropsychologia.2019.107277.

[83] Ito TA, Larsen JT, Smith NK, et al. Negative information weighs more heavily on the brain: the negativity bias in evaluative categorizations. J. Pers. Soc. Psychol. [Internet]. 1998;75:887–900. Available from: http://dx.doi.org/10.1037//0022-3514.75.4.887.

[84] Tversky A, Kahneman D. Prospect theory: An analysis of decision under risk. Econometrica [Internet]. 1979;47:263–291. Available from: http://ecsocman.hse.ru/rubezh/msg/16767948.html.

[85] Guerini M, Staiano J. Deep feelings: A massive cross-lingual study on the relation between emotions and virality. Proceedings of the 24th International Conference on World Wide Web [Internet]. 2015. p. 299–305. Available from: https://dl.acm.org/doi/abs/10.1145/2740908.2743058?casa_token=C7oXo7NMxB4AAAAA:pU8cbaYXACPcrLT_dBX3L_rhzvYXgFrrKSNi7BQaEwI3y-cPEKaWAyx4xhZOTuFhAH05rk3XMj5e.

[86] Goldie D, Linick M, Jabbar H, et al. Using Bibliometric and Social Media Analyses to Explore the “Echo Chamber” Hypothesis. Educ. Policy [Internet]. 2014;28:281–305. Available from: https://doi.org/10.1177/0895904813515330.

[87] Bail CA, Argyle LP, Brown TW, et al. Exposure to opposing views on social media can increase political polarization. Proc. Natl. Acad. Sci. U. S. A. [Internet]. 2018;115:9216–9221. Available from: http://dx.doi.org/10.1073/pnas.1804840115.

[88] Stefan Wojcik And. Sizing Up Twitter Users [Internet]. Pew Research Center. 2019 [cited 2019 Oct 1]. Available from: https://www.pewresearch.org/internet/2019/04/24/sizing-up-twitter-users/.

[89] Ptáček T, Habernal I, Hong J. Sarcasm detection on czech and english twitter. Proceedings of COLING 2014, the 25th International Conference on Computational Linguistics: Technical Papers [Internet]. 2014. p. 213–223. Available from: https://www.aclweb.org/anthology/C14-1022.pdf.

[90] Masud MT, Mamun MA, Thapa K, et al. Unobtrusive monitoring of behavior and movement patterns to detect clinical depression severity level via smartphone. J. Biomed. Inform. [Internet]. 2020;103:103371. Available from: http://dx.doi.org/10.1016/j.jbi.2019.103371.

Tables and Figures

Table 1: Number of tweets about each method per year with Mann Kendall statistics showing trend in numbers over time

	Number of tweets													
Method	**2007**	**2008**	**2009**	**2010**	**2011**	**2012**	**2013**	**2014**	**2015**	**2016**	**2017**	**2018**	**2019**	**Tau (p-value)***
IUD	16	333	676	9369	18439	28018	26317	34099	33238	39508	29712	68742	108762	0.87 (4.36E-05)
LNG-IUD	0	7	199	476	574	1068	2004	1065	1008	1268	1672	2480	3699	0.82 (0.000121)
Copper IUD	0	6	133	512	789	1977	1630	1306	1438	2146	2999	3991	7195	0.87 (4.36E-05)
Implant	0	20	452	1485	5120	6957	6956	7009	7529	8708	10967	15467	23445	0.97 (4.77E-06)
LARC Combined	16	346	1008	10357	19802	31063	29951	36470	35684	42922	34383	75213	119656	0.87 (4.36E-05)
Pill	14	120	1478	5262	8202	10301	8343	6066	5971	6817	9656	15516	19822	0.69 (0.001)
Patch	2	89	470	1934	3969	2317	1555	1318	758	614	797	945	1745	0.18 (0.428)
Ring	0	50	1556	4105	6279	10587	5531	7731	5089	6928	4396	5200	7195	0.33 (0.15)
Shot	3	54	1551	4729	12517	20516	17987	13838	10628	10260	9442	11751	15749	0.39 (0.08)
SARC Combined	19	313	5055	16030	30967	43721	33416	28953	22446	24619	24291	33412	44511	0.51 (0.02)
All methods	35	679	6515	27872	55889	81741	70323	72432	65659	76249	69641	124092	187612	N/A

* Mann Kendall test for monotonic trend

Table 2: Sensitivity and specificity of automated sentiment analysis in detection of sentiment toward the contraceptive method mentioned in a tweet, based on manual gold standard sentiment analysis

Sentiment	**Statistic**	**Amazon Comprehend Confidence Score (%)**			
		All	**80**	**90**	**95**
Positive	**Sensitivity (%)**	36	43	49	67
	Specificity (%)	85	91	91	92
Negative	**Sensitivity (%)**	66	82	84	83
	Specificity (%)	60	65	67	71
Neutral	**Sensitivity (%)**	64	76	77	74
	Specificity (%)	63	76	81	88
Mixed	**Sensitivity (%)**	14	0	0	0
	Specificity (%)	96	97	98	98

Table 3: Numbers of positive, negative and neutral tweets about each contraceptive method

Method	Total tweets	Number of tweets with ≥95% confident sentiment (n, %)*	Positive tweets (n, %)†	Negative tweets (n, %)†	Neutral tweets (n, %)†	P (positive vs. negative) ∮	P (positive vs. neutral) ∮	P (negative vs. neutral) ∮
IUD	280037	60277 (21.52%)	10602 (17.59%)	25039 (41.54%)	20314 (33.7%)	2.2e-16	2.2e-16	2.2e-16
LNG-IUD	11500	3729 (32.43%)	920 (24.67%)	590 (15.82%)	2071 (55.54%)	2.2e-16	2.2e-16	2.2e-16
Copper IUD	17577	4580 (26.06%)	1391 (30.37%)	1492 (32.58%)	1371 (29.93%)	0.06	0.70	0.02
Implant	76356	22724 (29.76%)	5026 (22.12%)	8628 (37.97%)	8015 (35.27%)	2.2e-16	2.2e-16	2.02e-06
LARC Combined	385470	91310 (23.69%)	17939 (19.65%)	35749 (39.15%)	31771 (34.79%)	2.2e-16	2.2e-16	2.2e-16
Pill	90836	20848 (22.95%)	1679 (8.05%)	5670 (27.2%)	12792 (61.36%)	2.2e-16	2.2e-16	2.2e-16
Patch	14568	4586 (31.48%)	335 (7.3%)	1455 (31.73%)	2663 (58.07%)	2.2e-16	2.2e-16	2.2e-16
Ring	56283	13389 (23.79%)	1928 (14.4%)	3660 (27.34%)	6353 (47.45%)	2.2e-16	2.2e-16	2.2e-16
Shot	117907	30580 (25.94%)	3144 (10.28%)	18805 (61.49%)	7698 (25.17%)	2.2e-16	2.2e-16	2.2e-16
SARC Combined	279594	69403 (24.82%)	7086 (10.21%)	29590 (42.64%)	29506 (42.51%)	2.2e-16	2.2e-16	0.73
All methods	665064	160713 (24.17%)	25025 (15.57%)	65339 (40.66%)	61277 (38.13%)	2.2e-16	2.2e-16	2.2e-16

* % of all tweets mentioning method

† % of tweets with ≥95% confident sentiment; note that percentages do not add up to 100 because mixed tweets were excluded from analysis on the basis of Amazon Comprehend's 0% sensitivity for detecting mixed emotion based on our manual gold standard analysis.

∮ Chi Square test of proportions for the numbers of positive vs. negative; positive vs. neutral; and negative vs. neutral tweets mentioning each method and method class.

Figure 1: Reversible, prescription contraceptive methods (A) and Percent of women using different contraceptive methods in the United States based on estimated from the National Survey of Family Growth, 2006 - 2017 (B)

A **B**

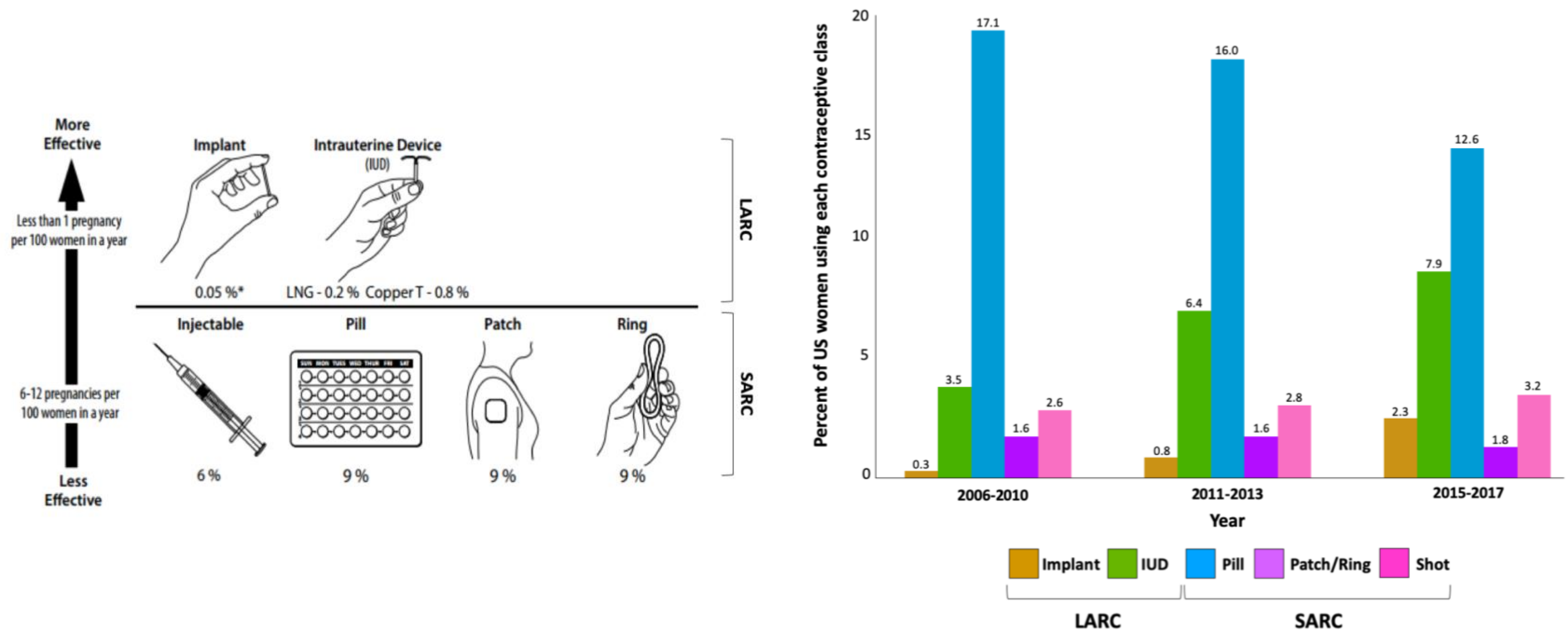


Figure 1 Legend: (A) Reversible, prescription contraceptive methods, adapted from https://www.cdc.gov/reproductivehealth/UnintendedPregnancy/PDF/effectiveness_of_contraceptive_methods.pdf. *Percentages indicate Pearl Index, i.e. number out of 100 women who experienced unintended pregnancy within the first year of typical use of each contraceptive method. **(B)** Percent of women using different contraceptive methods in the United States based on National Center for Health Statistics data, 2006 - 2017. 2006-2010 data are based on a sample of 12,279 women interviewed in 2006–2010; 2011-2013 estimates are based on data from the 5,601 women in the female respondent file of the 2011–2013 National Survey of Family Growth (NSFG); 2015-2017 estimates are based on data from the 5,554 women in the female respondent file of the 2015–2017 NSFG. No data were available for 2014 or 2018-2019. For the 2006-2010 data collection period, the implant, lunelle (1 month

injectable) and patch were grouped together; for the purposes of this graphic, it was assumed that 1/3 of this group used each method.

Figure 2: Tweet filtering and removal of irrelevant text

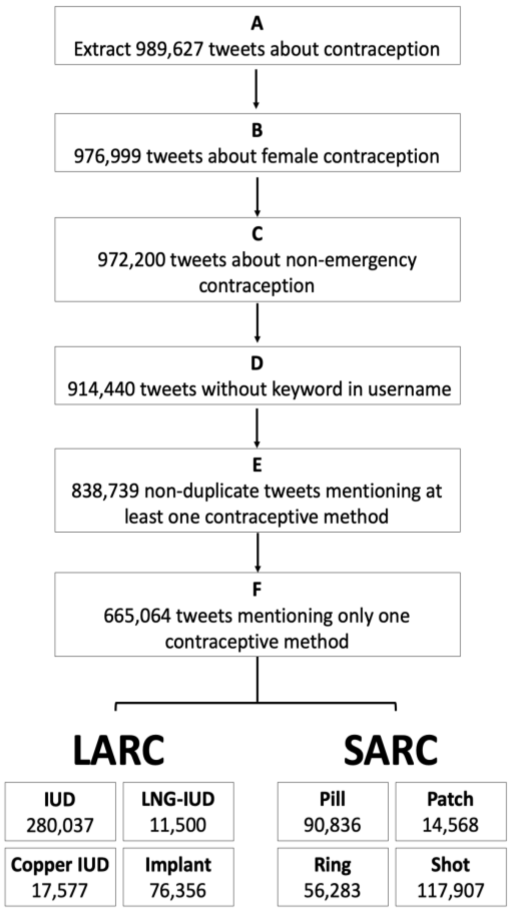


Figure 3: Workflow of data extraction, processing and analysis

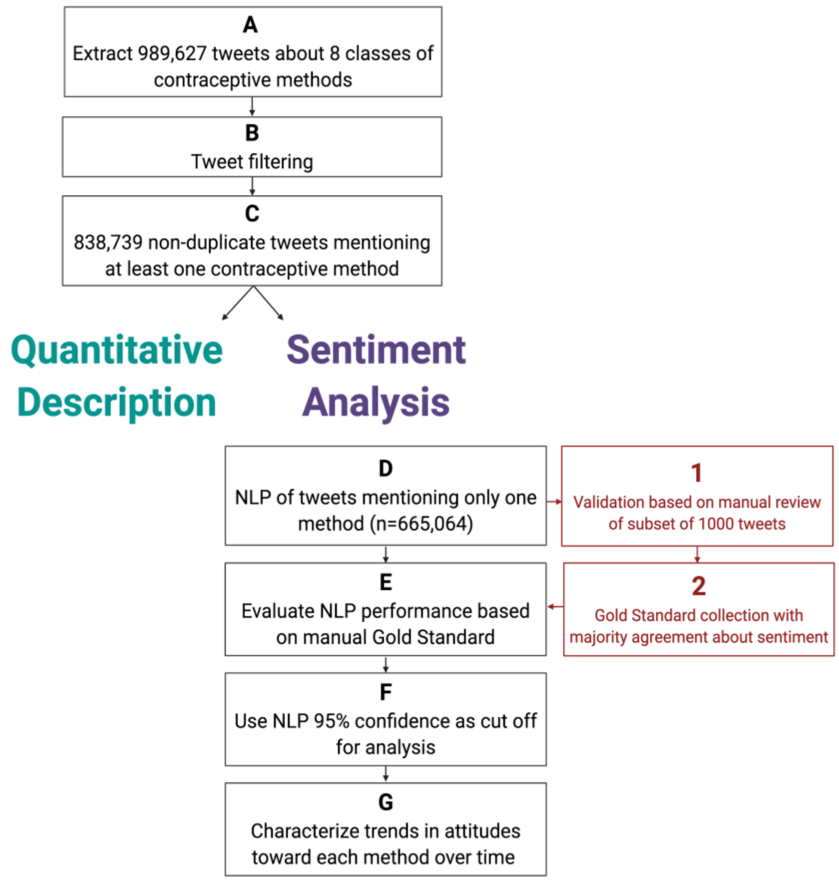


Figure 2 Legend: Workflow of tweet filtering and removal of irrelevant text **A.** Tweets were harvested using Python GetOldTweets3. 8 classes of contraception included IUD, Copper IUD, LNG-IUD, Implant, Pill, Patch, Ring, and Shot; see **Supplementary Table 1** for list of keywords used to harvest tweets; **A-B:** Remove 12,628 tweets containing the phrases “male contraception”, “male contraceptive”, and “male birth control”; **B-C:** Remove 4,799 tweets containing the phrases “emergency contraception”, “emergency contraceptive”, and “emergency birth control”; **C-D:** Remove 57,760 tweets where username contained a keyword (e.g. username @NuvaRingLawyer); see **Supplementary Table 1** for complete list of keywords; **D-E:** Remove 75,701 duplicate tweets, keeping the earliest one; **E-F:** Remove 173,675 tweets mentioning more than one contraceptive method to enable automated sentiment analysis.

Figure 3 Legend: General workflow of data extraction, processing, and analysis. **A.** Tweets were harvested using Python GetOldTweets3. 8 classes of contraception included IUD, Copper IUD, LNG-IUD, Implant, Pill, Patch, Ring, and Shot; see **Supplementary Table 1** for list of keywords used to harvest tweets. **B.** Removal of tweets about emergency contraception, male contraception, tweets harvested because the twitter username contains a keyword (e.g. @NivaRingLawyer), and duplicate tweets (see **Figure 3**). Tweet description included quantifying the number of tweets per contraception class, tweets per year, tweets per contraception class per year (see **Figure 4**), number of users, and number of tweets per user. Sentiment analysis was conducted to answer the question: is each tweet mentioning a single contraceptive method (n=665,064) positive, negative, neutral, or mixed emotion? **D.** Amazon Comprehend Sentiment Analysis API was used for automated sentiment analysis. **1.** Manual analysis of random sample of 973 tweets by two groups of five reviewers; interrater agreement calculated with a Kappa statistic (see **Supplementary Figure 1**). **2.** 889 tweets with agreement of ≥ 3 out of 5 reviewers make up "Gold Standard" sentiment analysis. **E.** We calculated the sensitivity and specificity of Amazon Comprehend for detection of positive, neutral, and negative sentiment regarding the contraceptive method mentioned in a given tweet based on manual Gold Standard from (**1-2**). **F.** Optimal sensitivity and specificity for correct tweet sentiment occurred at Amazon Comprehend’s 95% confidence score (see **Table 2**). **G**. Analysis of trends of number of positive, neutral and negative tweets about each contraceptive method between 2006 and 2019 (see **Figure 5**).

Figure 4: Number of tweets mentioning each class of contraception since 2006, in aggregate (A), per year (B), and per year, adjusted for the total number of mentions of contraception per year (C)

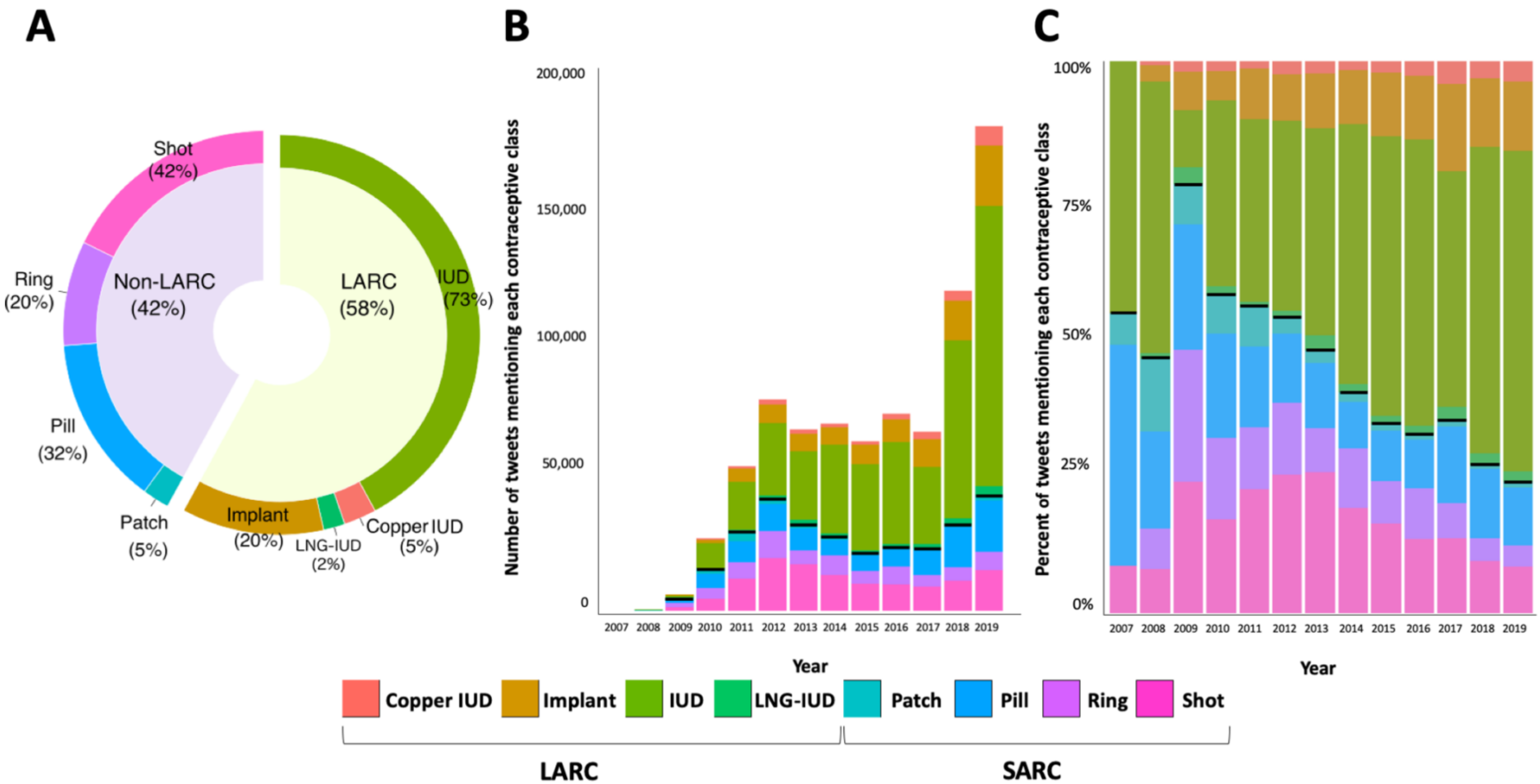


Figure 4 Legend: Number of filtered, non-duplicate tweets (Step **E** in **Figure 2**) mentioning each class of contraception since 2006, in aggregate (**A**), per year (**B**), and per year, adjusted for the total number of mentions of contraception per year (**C**). Black bars in (**B**) and (**C**) represent division between SARC (below) and LARC (above). Tweets mentioning multiple contraceptive methods count toward each method class once, meaning that a single tweet can contribute to multiple categories.

Figure 5: Trends in the proportion and number of positive, negative and neutral tweets about each contraceptive class per year since 2006

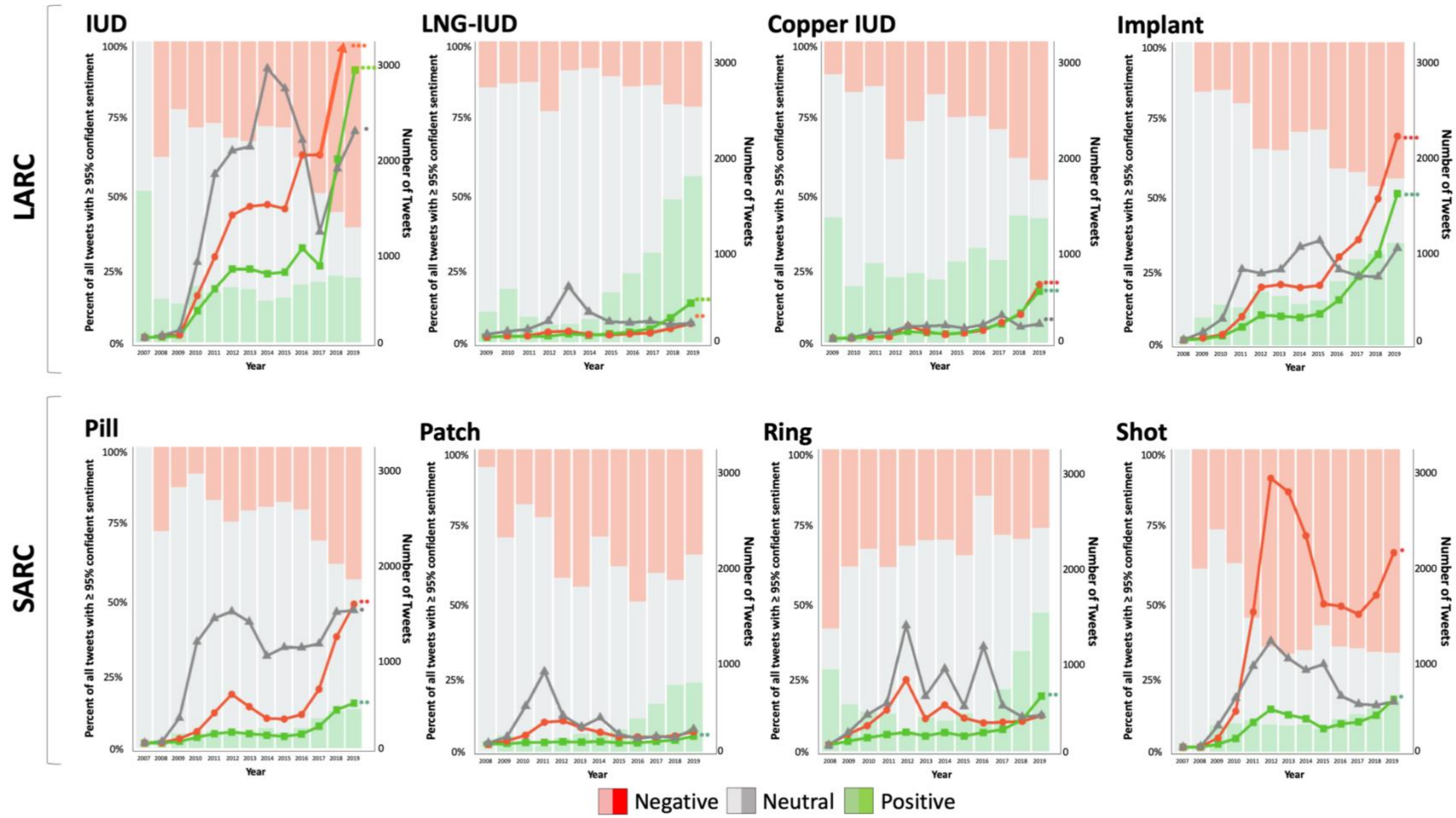


Figure 5 Legend: Trends in the proportion and number of positive, negative and neutral tweets about each contraceptive class per year since 2006. Bars represent proportions and lines represent numbers of tweets in each sentiment category. Mixed emotion

tweets were excluded due to NLP algorithm's 0% sensitivity in detecting mixed emotions toward a contraceptive method based on our gold standard manual sentiment analysis. Mann Kendall Tests were used to test for monotonic trends in number of tweets, with * denoting $p<0.05$, ** $p<0.005$, and *** $p<0.0005$. For tweets mentioning the IUD, the number of negative tweets in 2018 and 2019 were 5158 and 8654, respectively.

Appendix

Supplementary Table 1: Complete list of Twitter search keywords used and not used to harvest tweets about contraceptive methods

Contraceptive method	**Search keywords**		
Intrauterine device (IUD)	*Terms used*	*Number of Tweets*	*Terms not used*
	Intrauterine device IUD Intrauterine system Uterine jewelry **Total** pre-filtering* **Total** without male BC, EC, duplicates, or keyword in username	3978 483085 819 6 **483402** **397229**	IUS † Underbling †
Copper IUD	*Terms used*		*Terms not used*
	Copper-T Cu-IUD Paragard Paraguard Copper IUD Copper IUS Copper Intrauterine device Copper T birth control Copper T contraceptive **Total** pre-filtering* **Total** without male BC, EC, duplicates, or keyword in username	337 39 9163 5630 12083 2 249 6 3 **25800** **24122**	Copper T † Cu T † Cu-T † Copper Intrauterine System∮ Copper T contraception∮
Levonorgestrel IUD	*Terms used*		*Terms not used*
	Hormonal IUD Hormone IUD Hormonal IUS Hormone IUS Hormonal Intrauterine device Hormone Intrauterine device Hormonal Intrauterine system Progesterone IUD Progesterone IUS Progestin IUD Levonorgestrel IUD Levonorgestrel IUS	3298 219 35 4 49 2 8 90 7 101 156 17	Mirena † Skyla † Skyla IUS∮ Liletta † Liletta IUS∮ Kyleena † Kyleena Intrauterine Device∮ Kyleena Intrauterine System∮ Progesterone Intrauterine

	Levonorgestrel Intrauterine device Levonorgestrel Intrauterine system LNG-IUD LNG-IUS LNG IUD LNG IUS Mirena IUD Mirena IUS Mirena Intrauterine device Mirena Intrauterine system Skyla IUD Skyla Intrauterine device Skyla Intrauterine system Liletta IUD Liletta Intrauterine device Liletta Intrauterine system Kyleena IUD Kyleena IUS **Total** pre-filtering* **Total** without male BC, EC, duplicates, or keyword in username	62 160 36 233 42 78 12004 218 25 11 784 2 5 284 5 3 457 1 **16935** **15520**	device§ Progesterone Intrauterine system§ Progestin Intrauterine device§ Progestin Intrauterine system§ Progestin IUS§ Hormone Intrauterine system§
Implant	*Terms used*		*Terms not used*
	Nexplanon Implanon Arm Implant BC Implant B.C. Implant Birth Control Implant Contraception Implant Contraceptive Implant Birth Control Rod Contraception Rod Contraceptive Rod **Total** pre-filtering* **Total** without male BC, EC, duplicates, or keyword in username	44249 31787 8591 1952 43 13742 366 11731 590 19 89 **100552** **94131**	The rod † The implant †
Pill	*Terms used*		*Terms not used*
	OC pill O.C. pill BC pill B.C. pill Birth control pill Contraceptive pill Oral contraceptive Oral contraception Oral birth control	131 4 17198 675 32125 63767 25789 8262 2028	The pill † OCP † BCP † CHCP † COCP † Placebo period † Fake period † Faux period †

	Combined hormonal contraceptive CHC pill COC pill Combined BCP Combined OCP Combined B.C.P. Progestin only pill Progestin OCP **Total** pre-filtering* **Total** without male BC, EC, duplicates, or keyword in username	98 2 39 13 44 1 314 1 **125934** **99538**	Placebo week † C.H.C. pill** C.O.C. pill** Combined O.C.P. ∮ Progestin BCP ∮ Progestin B.C.P. ∮ Progestin O.C.P. ∮ Brand names (Alesse, Apri, Aranelle, Aviane, Enpresse, Estrostep, Lessina, Levlen, Levlite, Levora, Loestrin, Mircette, Natazia, Nordette, Lo/Orval, Ortho-Novum, Ortho Tri-Cyclen, Yasmin, Yaz, Seasonique, Seasonale, Jolessa, Qualsense, Lybrel, Camila, Errin, Jolivette, Micronor) †
Patch	*Terms used*		*Terms not used*
	BC patch Birth control patch Contraceptive patch Contraception patch Ortho evra Xulane **Total** pre-filtering* **Total** without male BC, EC, duplicates, or keyword in username	1957 13362 240 3175 3191 610 **19656** **16513**	The patch † B.C. patch †
Ring	*Terms used*		*Terms not used*
	Nuva ring Nuvaring Nuva-ring Vaginal ring Contraceptive ring Contraception ring Birth control ring Annovera **Total** pre-filtering* **Total** without male BC, EC, duplicates, or keyword in username	37068 22319 29400 15909 1078 194 1313 314 **72886** **64663**	The ring † Nuveen † B.C. ring † BC ring †
Shot	*Terms used*		*Terms not used*
	Birth control shot	35010	The shot †

	Contraceptive shot Contraception shot Birth control injection Contraceptive injection Contraception injection Depo provera Depo-provera Depo shot Depo injection Depo BC Depo birth control Depo contraception Depo contraceptive DMPA Medroxyprogesterone acetate Depot provera Depot shot Depot injection Depot birth control Depot contraception Depot contraceptive **Total** pre-filtering* **Total** without male BC, EC, duplicates, or keyword in username	800 74 2032 711 594 22270 21829 72423 2559 525 1160 39 89 5137 883 106 1353 988 16 31 46 **144462** **129050**	Depo B.C. † Depot B.C. † Depot BC † Depo † BC shot † B.C. shot †

BC: Birth Control; EC: Emergency Contraception.

***** Initial non-duplicate tweets, i.e. total harvested tweets in method class after removing duplicate tweets. For example, “loving my nexplanon contraceptive implant” was harvested twice because it contains two search phrases (“nexplanon” and “contraceptive implant”); only one copy counted toward the total initial tweet count.

† Keywords did not reliably return tweets related to birth control.

∮ Keywords returned 0 tweets.

Supplementary Table 2: Keywords used to distinguish between Birth Control Categories.

Category	Keyword
IUD	IUD Intrauterine Device Intrauterine System Uterine jewelry
Copper IUD	Copper Copper-T Copper-IUD Cu-IUD Paragard Paraguard Copper IUD Copper IUS Copper Intrauterine device Copper Intrauterine system Copper T birth control Copper T contraception Copper T contraceptive Non-hormonal
LNG-IUD	Hormonal IUD Hormone IUD Hormonal IUS Hormone IUS Levonorgestrel IUD Levonorgestrel IUS Levonorgestrel Intrauterine device Levonorgestrel Intrauterine system Progesterone IUD Progesterone IUS Progestin IUD LNG IUD LNG IUS LNG-IUD LNG-IUS Mirena IUD Mirena IUS Mirena Intrauterine device Mirena Intrauterine system Skyla IUD Skyla IUS Skyla Intrauterine device Skyla Intrauterine system Liletta IUD Liletta IUS Liletta Intrauterine device Liletta Intrauterine system Kyleena IUD Kyleena IUS Kyleena Intrauterine device Kyleena Intrauterine system

	Mirena Kyleena Skyla Liletta
Implant	Nexplanon Implanon Arm implant BC implant B.C. implant Birth control implant Contraceptive implant Contraception implant Implant
Shot	Birth control shot Contraceptive shot Contraception shot Birth control injection Contraceptive injection Contraception injection Depo provera Depo shot Depo injection Depo BC Depo birth control Depo contraception Depo contraceptive DMPA Medroxyprogesterone acetate Depot provera Depot shot Depot injection Depot bc Depot birth control Depot contraception Depot contraceptive Shot Depo Depot Injection
Pill	Pill OC pill B.C. pill BC pill Birth control pill Contraceptive pill Oral contraceptive Oral contraception Oral birth control Combined hormonal contraceptive CHC pill COC pill Combined BCP

	Combined OCP Progestin only pill Progestin BCP Progestin OCP
Ring	Nuva ring Nuvaring BC ring Vaginal ring Contraceptive ring Contraception ring Birth control ring Ring Nuveen Annovera
Patch	Patch BC patch Birth control patch Contraceptive patch Contraception patch Ortho evra Xulane

Supplementary Table 3: Number of tweets in each contraceptive method class at each stage of filtering

Method Class	Initial Tweets* (n, %) ^	Tweets not mentioning male contraception (n, %) †	Tweets not mentioning emergency contraception (n, %) †	Tweets where username does not contain a keyword (n, %) †	Tweets without duplicate text (e.g. retweets) (n, %) †	Tweets mentioning only 1 category (n, %) ∮
IUD	483402 (48.85)	483158 (99.95)	481290 (99.56)	427515 (88.44)	397229 (82.17)	280037 (70.50)
Copper IUD	25800 (2.61)	25795 (99.98)	25373 (98.34)	24923 (96.60)	24122 (93.50)	17577 (72.87)
LNG-IUD	16935 (1.71)	16932 (99.98)	16923 (99.93)	16751 (98.91)	15520 (91.64)	11500 (74.10)
Implant	100552 (10.16)	100514 (99.96)	100498 (99.95)	100495 (99.94)	94115 (93.60)	76356 (81.13)
Pill	125934 (12.73)	116562 (92.56)	114123 (90.62)	113887 (90.43)	97568 (77.48)	90836 (93.10)
Patch	19656 (1.99)	19583 (99.63)	19579 (99.61)	19569 (99.56)	16513 (84.01)	14568 (88.22)
Ring	72886 (7.36)	72868 (99.98)	72852 (99.95)	72261 (99.14)	64647 (88.70)	56283 (87.06)
Shot	144462 (14.60)	141587 (98.01)	141562 (97.99)	139039 (96.25)	129025 (89.31)	117907 (91.38)
Total	989627 (100)	976999 (98.72)	972200 (98.24)	914440 (92.40)	838739 (84.75)	NA

***** Initial non-duplicate tweets (i.e. total harvested tweets in method class after removing duplicate tweets; e.g. "loving my nexplanon contraceptive implant" was harvested twice because it contains two search phrases ("nexplanon" and "contraceptive implant"; only one copy counted toward the total initial tweet count)

^ % of all harvested tweets

† % of tweets in contraceptive class

∮ % of filtered, non-duplicate tweets (i.e. the column "Tweets without duplicate text (e.g. retweets)")

Supplementary Table 4: Sentiment analysis results for all methods combined

Year	Total tweets	Number of tweets with ≥95% confident sentiment (n, %)*	Positive tweets (n, %)†	Negative tweets (n, %)†	Neutral tweets (n, %)†	Mixed tweets (n, %)†
2007	31	5 (16.13%)	1 (20%)	0 (0%)	4 (80%)	0 (0%)
2008	552	120 (21.74%)	11 (9.17%)	39 (32.5%)	63 (52.5%)	7 (5.83%)
2009	5890	1554 (26.38%)	145 (9.33%)	377 (24.26%)	962 (61.9%)	70 (4.5%)
2010	24099	6006 (24.92%)	646 (10.76%)	1410 (23.48%)	3656 (60.87%)	294 (4.9%)
2011	48441	11841 (24.44%)	1250 (10.56%)	3686 (31.13%)	6351 (53.64%)	554 (4.68%)
2012	69507	16903 (24.32%)	1852 (10.96%)	6752 (39.95%)	7459 (44.13%)	840 (4.97%)
2013	59175	15020 (25.38%)	1726 (11.49%)	5965 (39.71%)	6713 (44.69%)	616 (4.1%)
2014	59339	15013 (25.3%)	1621 (10.8%)	5366 (35.74%)	7454 (49.65%)	572 (3.81%)
2015	51962	13125 (25.26%)	1526 (11.63%)	4393 (33.47%)	6720 (51.2%)	486 (3.7%)
2016	61690	14333 (23.23%)	2137 (14.91%)	5320 (37.12%)	6112 (42.64%)	764 (5.33%)
2017	56003	13503 (24.11%)	2444 (18.1%)	5811 (43.03%)	4454 (32.99%)	794 (5.88%)
2018	92630	21577 (23.29%)	4509 (20.9%)	10346 (47.95%)	5185 (24.03%)	1537 (7.12%)
2019	135745	31713 (23.36%)	7157 (22.57%)	15874 (50.06%)	6144 (19.37%)	2538 (8%)
Total	665064	160713 (24.17%)	25025 (15.57%)	65339 (40.66%)	61277 (38.13%)	9072 (5.64%)

* % of all tweets in that year

† % of tweets with ≥95% confident sentiment

Supplementary Table 5: Sentiment analysis results, all LARC methods combined

Year	Total tweets	Number of tweets with ≥95% confident sentiment (n, %)*	Positive tweets (n, %)†	Negative tweets (n, %)†	Neutral tweets (n, %)†	Mixed tweets (n, %)†
2007	13	2 (15.38%)	1 (50%)	0 (0%)	1 (50%)	0 (0%)
2008	296	56 (18.92%)	7 (12.5%)	19 (33.93%)	25 (44.64%)	5 (8.93%)
2009	1172	325 (27.73%)	39 (12%)	61 (18.77%)	214 (65.85%)	11 (3.38%)
2010	9060	2242 (24.75%)	380 (16.95%)	557 (24.84%)	1205 (53.75%)	100 (4.46%)
2011	19740	4886 (24.75%)	728 (14.9%)	1196 (24.48%)	2767 (56.63%)	195 (3.99%)
2012	28332	6742 (23.8%)	1130 (16.76%)	2154 (31.95%)	3143 (46.62%)	315 (4.67%)
2013	27598	7247 (26.26%)	1132 (15.62%)	2225 (30.7%)	3621 (49.97%)	269 (3.71%)
2014	32467	7967 (24.54%)	1037 (13.02%)	2143 (26.9%)	4470 (56.11%)	317 (3.98%)
2015	31356	7748 (24.71%)	1126 (14.53%)	2136 (27.57%)	4174 (53.87%)	312 (4.03%)
2016	38995	8515 (21.84%)	1615 (18.97%)	3094 (36.34%)	3306 (38.83%)	500 (5.87%)
2017	34144	8030 (23.52%)	1766 (21.99%)	3384 (42.14%)	2335 (29.08%)	545 (6.79%)
2018	63754	14643 (22.97%)	3447 (23.54%)	7100 (48.49%)	2861 (19.54%)	1235 (8.43%)
2019	98543	22907 (23.25%)	5531 (24.15%)	11680 (50.99%)	3649 (15.93%)	2047 (8.94%)
Total	385470	91310 (23.69%)	17939 (19.65%)	35749 (39.15%)	31771 (34.79%)	5851 (6.41%)

* % of all tweets in that year

† % of tweets with ≥95% confident sentiment

Supplementary Table 6: Sentiment analysis results, all SARC methods combined

Year	Total tweets	Number of tweets with ≥95% confident sentiment (n, %)*	Positive tweets (n, %)†	Negative tweets (n, %)†	Neutral tweets (n, %)†	Mixed tweets (n, %)†
2007	18	3 (16.67%)	0 (0%)	0 (0%)	3 (100%)	0 (0%)
2008	256	64 (25%)	4 (6.25%)	20 (31.25%)	38 (59.38%)	2 (3.12%)
2009	4718	1229 (26.05%)	106 (8.62%)	316 (25.71%)	748 (60.86%)	59 (4.8%)
2010	15039	3764 (25.03%)	266 (7.07%)	853 (22.66%)	2451 (65.12%)	194 (5.15%)
2011	28701	6955 (24.23%)	522 (7.51%)	2490 (35.8%)	3584 (51.53%)	359 (5.16%)
2012	41175	10161 (24.68%)	722 (7.11%)	4598 (45.25%)	4316 (42.48%)	525 (5.17%)
2013	31577	7773 (24.62%)	594 (7.64%)	3740 (48.12%)	3092 (39.78%)	347 (4.46%)
2014	26872	7046 (26.22%)	584 (8.29%)	3223 (45.74%)	2984 (42.35%)	255 (3.62%)
2015	20606	5377 (26.09%)	400 (7.44%)	2257 (41.98%)	2546 (47.35%)	174 (3.24%)
2016	22695	5818 (25.64%)	522 (8.97%)	2226 (38.26%)	2806 (48.23%)	264 (4.54%)
2017	21859	5473 (25.04%)	678 (12.39%)	2427 (44.34%)	2119 (38.72%)	249 (4.55%)
2018	28876	6934 (24.01%)	1062 (15.32%)	3246 (46.81%)	2324 (33.52%)	302 (4.36%)
2019	37202	8806 (23.67%)	1626 (18.46%)	4194 (47.63%)	2495 (28.33%)	491 (5.58%)
Total	279594	69403 (24.82%)	7086 (10.21%)	29590 (42.64%)	29506 (42.51%)	3221 (4.64%)

* % of all tweets in that year

† % of tweets with ≥95% confident sentiment

Supplementary Table 7: Sentiment analysis results, IUD

Year	Total tweets	Number of tweets with ≥95% confident sentiment (n, %)*	Positive tweets (n, %)†	Negative tweets (n, %)†	Neutral tweets (n, %)†	Mixed tweets (n, %)†
2007	13	2 (15.38%)	1 (50%)	0 (0%)	1 (50%)	0 (0%)
2008	265	52 (19.62%)	7 (13.46%)	19 (36.54%)	23 (44.23%)	3 (5.77%)
2009	509	131 (25.74%)	16 (12.21%)	29 (22.14%)	81 (61.83%)	5 (3.82%)
2010	7021	1692 (24.1%)	300 (17.73%)	466 (27.54%)	841 (49.7%)	85 (5.02%)
2011	14268	3417 (23.95%)	542 (15.86%)	900 (26.34%)	1820 (53.26%)	155 (4.54%)
2012	20051	4441 (22.15%)	762 (17.16%)	1364 (30.71%)	2082 (46.88%)	233 (5.25%)
2013	19026	4548 (23.9%)	762 (16.75%)	1461 (32.12%)	2128 (46.79%)	197 (4.33%)
2014	24814	5439 (21.92%)	710 (13.05%)	1482 (27.25%)	2998 (55.12%)	249 (4.58%)
2015	23265	5170 (22.22%)	729 (14.1%)	1435 (27.76%)	2775 (53.68%)	231 (4.47%)
2016	29167	5611 (19.24%)	997 (17.77%)	2035 (36.27%)	2201 (39.23%)	378 (6.74%)
2017	21518	4355 (20.24%)	799 (18.35%)	2036 (46.75%)	1181 (27.12%)	339 (7.78%)
2018	47036	9955 (21.16%)	1988 (19.97%)	5158 (51.81%)	1883 (18.92%)	926 (9.3%)
2019	73084	15464 (21.16%)	2989 (19.33%)	8654 (55.96%)	2300 (14.87%)	1521 (9.84%)
Total	280037	60277 (21.52%)	10602 (17.59%)	25039 (41.54%)	20314 (33.7%)	4322 (7.17%)

* % of all tweets in that year

† % of tweets with ≥95% confident sentiment

Supplementary Table 8: Sentiment analysis results, LNG-IUD

Year	Total tweets	Number of tweets with ≥95% confident sentiment (n, %)*	Positive tweets (n, %)†	Negative tweets (n, %)†	Neutral tweets (n, %)†	Mixed tweets (n, %)†
2008	6	1 (16.67%)	0 (0%)	0 (0%)	0 (0%)	1 (100%)
2009	174	53 (30.46%)	5 (9.43%)	8 (15.09%)	37 (69.81%)	3 (5.66%)
2010	381	106 (27.82%)	18 (16.98%)	15 (14.15%)	69 (65.09%)	4 (3.77%)
2011	481	123 (25.57%)	10 (8.13%)	17 (13.82%)	92 (74.8%)	4 (3.25%)
2012	890	267 (30%)	15 (5.62%)	62 (23.22%)	183 (68.54%)	7 (2.62%)
2013	1658	684 (41.25%)	41 (5.99%)	71 (10.38%)	569 (83.19%)	3 (0.44%)
2014	839	351 (41.84%)	27 (7.69%)	34 (9.69%)	287 (81.77%)	3 (0.85%)
2015	731	253 (34.61%)	41 (16.21%)	31 (12.25%)	178 (70.36%)	3 (1.19%)
2016	945	280 (29.63%)	62 (22.14%)	43 (15.36%)	167 (59.64%)	8 (2.86%)
2017	1235	355 (28.74%)	98 (27.61%)	51 (14.37%)	183 (51.55%)	23 (6.48%)
2018	1686	499 (29.6%)	219 (43.89%)	101 (20.24%)	145 (29.06%)	34 (6.81%)
2019	2474	757 (30.6%)	384 (50.73%)	157 (20.74%)	161 (21.27%)	55 (7.27%)
Total	11500	3729 (32.43%)	920 (24.67%)	590 (15.82%)	2071 (55.54%)	148 (3.97%

* % of all tweets in that year

† % of tweets with ≥95% confident sentiment

Supplementary Table 9: Sentiment analysis results, Copper IUD

Year	Total tweets	Number of tweets with ≥95% confident sentiment (n, %)*	Positive tweets (n, %)†	Negative tweets (n, %)†	Neutral tweets (n, %)†	Mixed tweets (n, %)†
2008	6	1 (16.67%)	0 (0%)	0 (0%)	0 (0%)	1 (100%)
2009	97	18 (18.56%)	7 (38.89%)	2 (11.11%)	8 (44.44%)	1 (5.56%)
2010	384	95 (24.74%)	17 (17.89%)	16 (16.84%)	58 (61.05%)	4 (4.21%)
2011	582	121 (20.79%)	30 (24.79%)	18 (14.88%)	67 (55.37%)	6 (4.96%)
2012	1473	381 (25.87%)	76 (19.95%)	140 (36.75%)	137 (35.96%)	28 (7.35%)
2013	1160	289 (24.91%)	63 (21.8%)	75 (25.95%)	138 (47.75%)	13 (4.5%)
2014	933	250 (26.8%)	50 (20%)	44 (17.6%)	146 (58.4%)	10 (4%)
2015	1023	260 (25.42%)	64 (24.62%)	62 (23.85%)	114 (43.85%)	20 (7.69%)
2016	1589	384 (24.17%)	110 (28.65%)	90 (23.44%)	154 (40.1%)	30 (7.81%)
2017	2316	645 (27.85%)	163 (25.27%)	179 (27.75%)	258 (40%)	45 (6.98%)
2018	2861	747 (26.11%)	284 (38.02%)	267 (35.74%)	129 (17.27%)	67 (8.97%)
2019	5153	1389 (26.96%)	527 (37.94%)	599 (43.12%)	162 (11.66%)	101 (7.27%)
Total	17577	4580 (26.06%)	1391 (30.37%)	1492 (32.58%)	1371 (29.93%)	326 (7.12%)

* % of all tweets in that year

† % of tweets with ≥95% confident sentiment

Supplementary Table 10: Sentiment analysis results, Implant

Year	Total tweets	Number of tweets with ≥95% confident sentiment (n, %)*	Positive tweets (n, %)†	Negative tweets (n, %)†	Neutral tweets (n, %)†	Mixed tweets (n, %)†
2008	19	2 (10.53%)	0 (0%)	0 (0%)	2 (100%)	0 (0%)
2009	392	123 (31.38%)	11 (8.94%)	22 (17.89%)	88 (71.54%)	2 (1.63%)
2010	1274	349 (27.39%)	45 (12.89%)	60 (17.19%)	237 (67.91%)	7 (2.01%)
2011	4409	1225 (27.78%)	146 (11.92%)	261 (21.31%)	788 (64.33%)	30 (2.45%)
2012	5918	1653 (27.93%)	277 (16.76%)	588 (35.57%)	741 (44.83%)	47 (2.84%)
2013	5754	1726 (30%)	266 (15.41%)	618 (35.81%)	786 (45.54%)	56 (3.24%)
2014	5881	1927 (32.77%)	250 (12.97%)	583 (30.25%)	1039 (53.92%)	55 (2.85%)
2015	6337	2065 (32.59%)	292 (14.14%)	608 (29.44%)	1107 (53.61%)	58 (2.81%)
2016	7294	2240 (30.71%)	446 (19.91%)	926 (41.34%)	784 (35%)	84 (3.75%)
2017	9075	2675 (29.48%)	706 (26.39%)	1118 (41.79%)	713 (26.65%)	138 (5.16%)
2018	12171	3442 (28.28%)	956 (27.77%)	1574 (45.73%)	704 (20.45%)	208 (6.04%)
2019	17832	5297 (29.71%)	1631 (30.79%)	2270 (42.85%)	1026 (19.37%)	370 (6.99%)
Total	76356	22724 (29.76%)	5026 (22.12%)	8628 (37.97%)	8015 (35.27%)	1055 (4.64%)

* % of all tweets in that year

† % of tweets with ≥95% confident sentiment

Supplementary Table 11: Sentiment analysis results, Pill

Year	Total tweets	Number of tweets with ≥95% confident sentiment (n, %)*	Positive tweets (n, %)†	Negative tweets (n, %)†	Neutral tweets (n, %)†	Mixed tweets (n, %)†
2007	14	2 (14.29%)	0 (0%)	0 (0%)	2 (100%)	0 (0%)
2008	114	21 (18.42%)	0 (0%)	6 (28.57%)	15 (71.43%)	0 (0%)
2009	1423	354 (24.88%)	17 (4.8%)	49 (13.84%)	280 (79.1%)	8 (2.26%)
2010	5092	1354 (26.59%)	65 (4.8%)	129 (9.53%)	1129 (83.38%)	31 (2.29%)
2011	7686	1906 (24.8%)	103 (5.4%)	337 (17.68%)	1392 (73.03%)	74 (3.88%)
2012	9770	2228 (22.8%)	124 (5.57%)	545 (24.46%)	1468 (65.89%)	91 (4.08%)
2013	7941	1920 (24.18%)	103 (5.36%)	406 (21.15%)	1353 (70.47%)	58 (3.02%)
2014	5697	1381 (24.24%)	89 (6.44%)	276 (19.99%)	972 (70.38%)	44 (3.19%)
2015	5591	1444 (25.83%)	74 (5.12%)	268 (18.56%)	1068 (73.96%)	34 (2.35%)
2016	6404	1525 (23.81%)	99 (6.49%)	319 (20.92%)	1066 (69.9%)	41 (2.69%)
2017	9093	1971 (21.68%)	187 (9.49%)	601 (30.49%)	1107 (56.16%)	76 (3.86%)
2018	14197	3124 (22%)	371 (11.88%)	1186 (37.96%)	1461 (46.77%)	106 (3.39%)
2019	17814	3618 (20.31%)	447 (12.35%)	1548 (42.79%)	1479 (40.88%)	144 (3.98%)
Total	90836	20848 (22.95%)	1679 (8.05%)	5670 (27.2%)	12792 (61.36%)	707 (3.39%)

* % of all tweets in that year

† % of tweets with ≥95% confident sentiment

Supplementary Table 12: Sentiment analysis results, Patch

Year	Total tweets	Number of tweets with ≥95% confident sentiment (n, %)*	Positive tweets (n, %)†	Negative tweets (n, %)†	Neutral tweets (n, %)†	Mixed tweets (n, %)†
2007	2	0 (0%)	0 (NaN%)	0 (NaN%)	0 (NaN%)	0 (NaN%)
2008	43	16 (37.21%)	0 (0%)	1 (6.25%)	15 (93.75%)	0 (0%)
2009	418	134 (32.06%)	7 (5.22%)	38 (28.36%)	84 (62.69%)	5 (3.73%)
2010	1749	562 (32.13%)	19 (3.38%)	102 (18.15%)	427 (75.98%)	14 (2.49%)
2011	3447	1099 (31.88%)	21 (1.91%)	247 (22.47%)	812 (73.89%)	19 (1.73%)
2012	2036	630 (30.94%)	28 (4.44%)	262 (41.59%)	322 (51.11%)	18 (2.86%)
2013	1434	419 (29.22%)	24 (5.73%)	185 (44.15%)	196 (46.78%)	14 (3.34%)
2014	1228	472 (38.44%)	31 (6.57%)	135 (28.6%)	297 (62.92%)	9 (1.91%)
2015	692	222 (32.08%)	16 (7.21%)	84 (37.84%)	116 (52.25%)	6 (2.7%)
2016	548	170 (31.02%)	18 (10.59%)	83 (48.82%)	63 (37.06%)	6 (3.53%)
2017	722	207 (28.67%)	32 (15.46%)	83 (40.1%)	87 (42.03%)	5 (2.42%)
2018	798	223 (27.94%)	47 (21.08%)	93 (41.7%)	74 (33.18%)	9 (4.04%)
2019	1451	432 (29.77%)	92 (21.3%)	142 (32.87%)	170 (39.35%)	28 (6.48%)
Total	14568	4586 (31.48%)	335 (7.3%)	1455 (31.73%)	2663 (58.07%)	133 (2.9%)

* % of all tweets in that year

† % of tweets with ≥95% confident sentiment

Supplementary Table 13: Sentiment analysis results, Ring

Year	Total tweets	Number of tweets with ≥95% confident sentiment (n, %)*	Positive tweets (n, %)†	Negative tweets (n, %)†	Neutral tweets (n, %)†	Mixed tweets (n, %)†
2008	49	17 (34.69%)	4 (23.53%)	9 (52.94%)	2 (11.76%)	2 (11.76%)
2009	1451	359 (24.74%)	49 (13.65%)	127 (35.38%)	143 (39.83%)	40 (11.14%)
2010	3829	768 (20.06%)	85 (11.07%)	221 (28.78%)	344 (44.79%)	118 (15.36%)
2011	5813	1175 (20.21%)	121 (10.3%)	396 (33.7%)	473 (40.26%)	185 (15.74%)
2012	9879	2501 (25.32%)	148 (5.92%)	732 (29.27%)	1340 (53.58%)	281 (11.24%)
2013	5147	1112 (21.6%)	106 (9.53%)	300 (26.98%)	552 (49.64%)	154 (13.85%)
2014	6873	1579 (22.97%)	146 (9.25%)	453 (28.69%)	856 (54.21%)	124 (7.85%)
2015	4356	935 (21.46%)	105 (11.23%)	306 (32.73%)	437 (46.74%)	87 (9.3%)
2016	6267	1594 (25.43%)	146 (9.16%)	252 (15.81%)	1105 (69.32%)	91 (5.71%)
2017	3646	977 (26.8%)	179 (18.32%)	260 (26.61%)	444 (45.45%)	94 (9.62%)
2018	3936	976 (24.8%)	286 (29.3%)	270 (27.66%)	320 (32.79%)	100 (10.25%)
2019	5037	1396 (27.71%)	553 (39.61%)	334 (23.93%)	337 (24.14%)	172 (12.32%)
Total	56283	13389 (23.79%)	1928 (14.4%)	3660 (27.34%)	6353 (47.45%)	1448 (10.81%)

* % of all tweets in that year

† % of tweets with ≥95% confident sentiment

Supplementary Table 14: Sentiment analysis results, Shot

Year	Total tweets	Number of tweets with ≥95% confident sentiment (n, %)*	Positive tweets (n, %)†	Negative tweets (n, %)†	Neutral tweets (n, %)†	Mixed tweets (n, %)†
2007	2	1 (50%)	0 (0%)	0 (0%)	1 (100%)	0 (0%)
2008	50	10 (20%)	0 (0%)	4 (40%)	6 (60%)	0 (0%)
2009	1426	382 (26.79%)	33 (8.64%)	102 (26.7%)	241 (63.09%)	6 (1.57%)
2010	4369	1080 (24.72%)	97 (8.98%)	401 (37.13%)	551 (51.02%)	31 (2.87%)
2011	11755	2775 (23.61%)	277 (9.98%)	1510 (54.41%)	907 (32.68%)	81 (2.92%)
2012	19490	4802 (24.64%)	422 (8.79%)	3059 (63.7%)	1186 (24.7%)	135 (2.81%)
2013	17055	4322 (25.34%)	361 (8.35%)	2849 (65.92%)	991 (22.93%)	121 (2.8%)
2014	13074	3614 (27.64%)	318 (8.8%)	2359 (65.27%)	859 (23.77%)	78 (2.16%)
2015	9967	2776 (27.85%)	205 (7.38%)	1599 (57.6%)	925 (33.32%)	47 (1.69%)
2016	9476	2529 (26.69%)	259 (10.24%)	1572 (62.16%)	572 (22.62%)	126 (4.98%)
2017	8398	2318 (27.6%)	280 (12.08%)	1483 (63.98%)	481 (20.75%)	74 (3.19%)
2018	9945	2611 (26.25%)	358 (13.71%)	1697 (64.99%)	469 (17.96%)	87 (3.33%)
2019	12900	3360 (26.05%)	534 (15.89%)	2170 (64.58%)	509 (15.15%)	147 (4.38%)
Total	117907	30580 (25.94%)	3144 (10.28%)	18805 (61.49%)	7698 (25.17%)	933 (3.05%)

* % of all tweets in that year

† % of tweets with ≥95% confident sentiment

Supplementary Figure 1: Workflow of manual sentiment analysis for validation of NLP algorithm

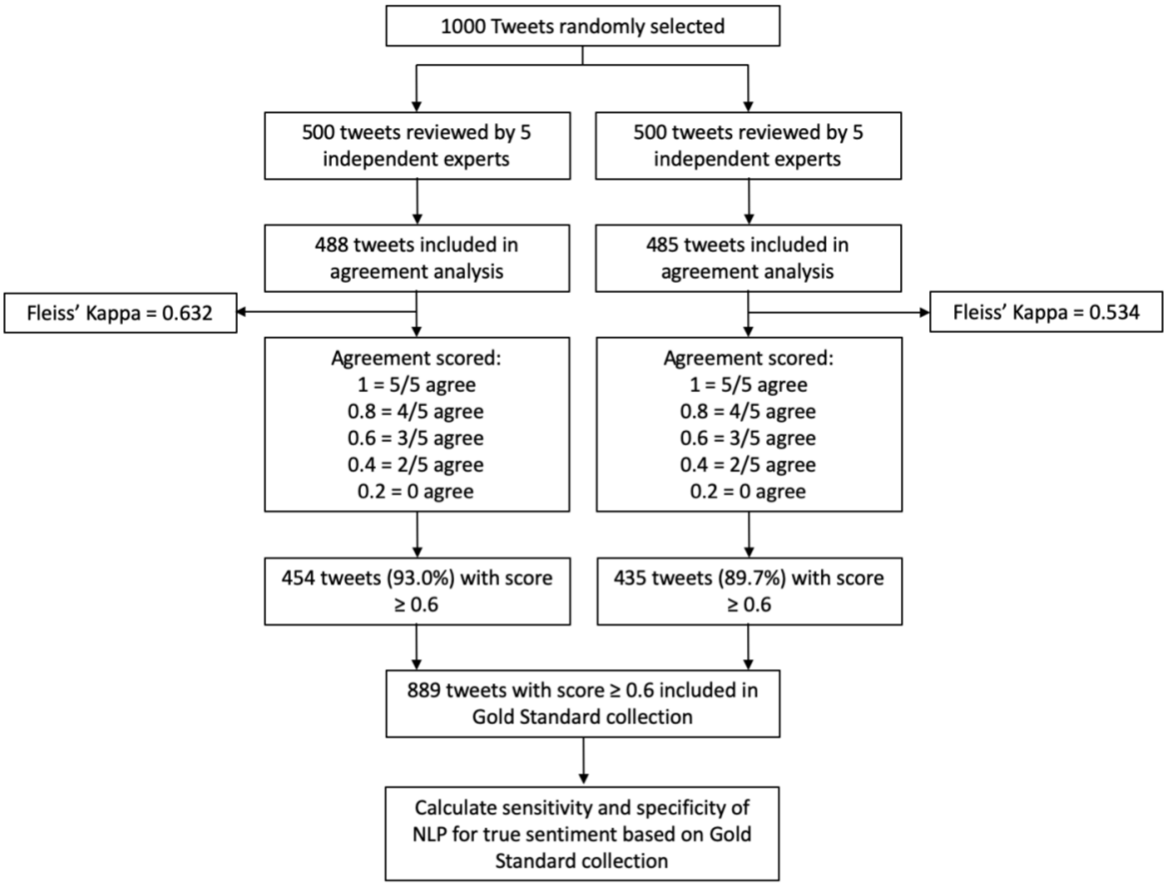


Figure Legend: Workflow of manual sentiment analysis for validation of NLP algorithm. 27 tweets were excluded from agreement analysis because in manual curation of tweets during review, we detected two classes of tweets, one related to male contraception and other related to emergency contraception, which were outside of the scope of this project. We thus decided to include an additional filter to exclude these classes of tweets from the entire collection (steps **B** and **C** in **Figure 3**). Tweets with majority (≥3 out of 5 reviewers) agreement on tweeter's sentiment toward contraceptive method mentioned in tweet were included in the Gold Standard collection, which was then used to calculate the sensitivity and specificity of the Amazon Comprehend Sentiment Analysis API.

Supplementary Document 1: Tweet Interpretation Guide

Introduction

Welcome to the 'Birth Control Attitudes on Twitter' research team! The aim of this project is to explore the attitudes toward different contraceptive methods (specifically reversible, prescription contraceptive methods resulting in <10 undesired pregnancies/year, i.e., the shot; oral contraceptives; the vaginal ring; the patch; the Copper IUD, Levonorgestrel IUDs; and Implanon/Nexplanon implants) as presented on Twitter since it was founded in 2006.

The broader purposes and implications of this work are to (1) better understand what is said on Twitter about contraceptive methods and how this has changed over the last 13 years because this has not been characterized at all in the literature, (2) better anticipate and address patients' concerns about specific birth control methods (based on the fact that twitter provides unfiltered information from patients compared to potentially biased information in surveys from researchers), and (3) be able to inform dissemination of accurate information about contraception via social media.

Methods

We have collected English language tweets that mention any one of the birth control methods listed above and are using computer programs to analyze the predominant emotions that can be detected in each individual tweet. By aggregating this data, we can look at the predominant sentiments toward different methods over time. For example, of the 99,712 tweets mentioning birth control pills since 2006 (when Twitter was founded), 8% are strongly positive and 27% are strongly negative.

To test how well the machine is working, we need to compare its performance to a human. This is where YOU come in! The ultimate goal is for each member of the 'human analysis team' to read and interpret the emotions in 500 tweets. More specific instructions for how to do this are detailed below. We will start with a set of 20 tweets as a trial run, and then we'll debrief to see how it's going. Once we've debriefed together, I will send each of you a google sheet containing 500 tweets that you'll read and interpret. You will be working blinded from one another, and then we'll compare your results to each other with a kappa statistic. The final aggregated results of this analysis will serve as the 'gold standard' sentiment analysis. Ultimately, we'll calculate the sensitivity and specificity for emotion detection of the computer program based on this gold standard generated by your human analysis.

The text that follows contains a guide for tweet interpretation. It is VERY IMPORTANT that you follow this guide CLOSELY. As you can imagine, interpreting the emotions in tweets could be very subjective and vary significantly between people. Our goal is to standardize it somewhat while still allowing for the subjectivity that makes us human.

Sentiment Analysis Instructions

Our goal is to determine each tweet's emotional valence, if any is present, with respect to the birth control method it mentions. Your task is to figure out whether each tweet is **positive, negative, neutral, or mixed *with respect to the method it mentions***, OR if it's a **false positive** (i.e. it isn't actually about birth control). Each tweet can only fall into one category. In simpler terms, when you read a tweet, ask yourself: "Is this person saying something good about this method, or something bad?" If the tweet doesn't have any emotional valence (e.g. "*ladies, what do you think of nexplanon*?"), mark it as neutral. If the tweet has mixed emotions (e.g. "*i didn't have weight gain or skin issues with the nexplanon arm implant but it fucked my period up. it's not terrible though, better than nothing*"), mark it as mixed. If it's not clear, it's always okay to say that it's mixed. A list of examples (straightforward and complex) is included below in **Tables 1 and 2** for further guidance.

Table 1: Straightforward tweets

Tweet	**Polarity**
this is why i'm so happy with my contraceptive arm implant. i get no period which is a big improvement from the constant bleeding, leaking and non existent cycle i used to have	Positive
i have nexplanon (arm implant) and i love it.	Positive
i would like to know more! had the arm implant and am having it removed because i hate it!	Negative
the arm implant terrifies me i've read so many horror stories	Negative
tell me about your experience with the birth control arm implant	Neutral
nexoplanon, the arm implant hehe	Neutral
i have the nexplanon arm implant thingy, and i haven't had any issues. the first week i was a lil crazy and had a headache, but i think it was my hormones in general leveling out post partum (i got it 6 weeks after delivery)	Mixed
isnt there another way? the arm implant? so you dont have to put metal in your twat thats a choice you made genius	Mixed
i think she has parkinsons bc pill rolling tremors are part of that disorder	False Positive

Much of the time, humans and the computer will agree. For example, "*f*ck you nexplanon.*" is interpreted by the computer as NEGATIVE, and probably most humans would agree. Another example: "*so glad i got my nexplanon implant today. if anyone is considering it i* *seriously recommend it.*" is interpreted by the computer as POSITIVE, and again, probably most humans would agree.

However, there are instances when the computer falls short. We want to know the **attitude toward the birth control method** mentioned in the tweet, which is sometimes different from the attitude of the tweet in general. A shortcoming of the computer is that it only detects the attitude of the tweet in general. For example, "*ugh, i just want nexplanon. but it's so expensive*." is interpreted by the computer as NEGATIVE. But we can tell that the tweeter actually seems to feel positively about Nexplanon; they just aren't able to afford it. Overall, the emotions in this tweet are mixed.

Sometimes, it's just not really clear if a tweet is saying anything positive or negative about the birth control method it mentions. For example, the tweet *"i had a friend that had an arm implant and it traveled and was messing up her hormones and she started losing her sight and bleeding non-stop. that's what's crazy about the human body, everyone experiences things differently, but it's great that we have so many options nowadays!"* is interpreted by the computer as POSITIVE, but it seems to me that this tweeter is expressing multiple emotions - they're sharing negative sentiments regarding their friend's experience with Nexplanon, but they seem to have a positive outlook overall, so ultimately, from a human perspective, this tweet is mixed.

Some tweets are replying to other twitter users, and include copied text from other tweets. Our code that 'cleans' the tweets and leaves us with only text that can be interpreted for sentiments can't distinguish that there are two voices in a tweet. For example, in the tweet "*i've had it for 3 years and i'm so scared right now. consider nexplanon! it goes in your arm instead,"* the tweeter has first copied the text of the tweet they're responding to, and then responded to it. This tweet is interpreted by the computer as NEGATIVE, but it sounds like the tweeter is responding to another person who is scared of their non-Nexplanon LARC and feels positively about Nexplanon. So ultimately (after this mental gymnastics), a human would probably consider this to be 'positive' about Nexplanon from the perspective of the tweeter.

Another shortcoming of the computer is that it can't detect sarcasm. For example, *"got my nexplanon rod taken out and they gave me no pain meds for the hole in my arm.. so that's good"* is interpreted by the computer as POSITIVE, but we as humans will probably recognize that it's sarcastic, and the true sentiment is negative.

Finally, a *small* number of the tweets we collected are not actually about birth control. This is because to collect tweets, we searched for keywords like "nexplanon" and "birth control pill." We also included abbreviations like "bc pill" which, most of the time, returns tweets about birth control pills, but every once in a while, returns a tweet like "i think she has parkinsons bc pill rolling tremors are part of that disorder". This would be a 'false positive.'

Table 2: Confusing tweets

Tweet	Computer	Human Comments
ugh, i just want nexplanon. but it's so expensive	Negative	we can tell that the tweeter actually seems to feel positively about Nexplanon; they just aren't able to afford it. Overall, the emotions in this tweet are mixed.
i had a friend that had an arm implant and it traveled and was messing up her hormones and she started losing her sight and bleeding non-stop. that's what's crazy about the human body, everyone experiences things differently, but it's great that we have so many options nowadays!	Positive	it seems to me that this tweeter is expressing multiple emotions - their friend's experience with Nexplanon was clearly bad, but they seem to have a positive outlook overall, so ultimately, from a human perspective, this tweet is mixed.
i've had it for 3 years and i'm so scared right now. consider nexplanon! it goes in your arm instead	Negative	it sounds like the tweeter is responding to another person who is scared of their non-Nexplanon LARC and feels positively about Nexplanon. So ultimately (after these mental gymnastics), a human would probably consider this to be 'positive' about Nexplanon from the perspective of the tweeter.
got my nexplanon rod taken out and they gave me no pain meds for the hole in my arm.. so that's good	Positive	This is SARCASM! The computer can't detect sarcasm, but we as humans do, and recognize that the true sentiment is negative.
i think she has parkinsons bc pill rolling tremors are part of that disorder	Neutral	False positive! This tweet is not actually about birth control

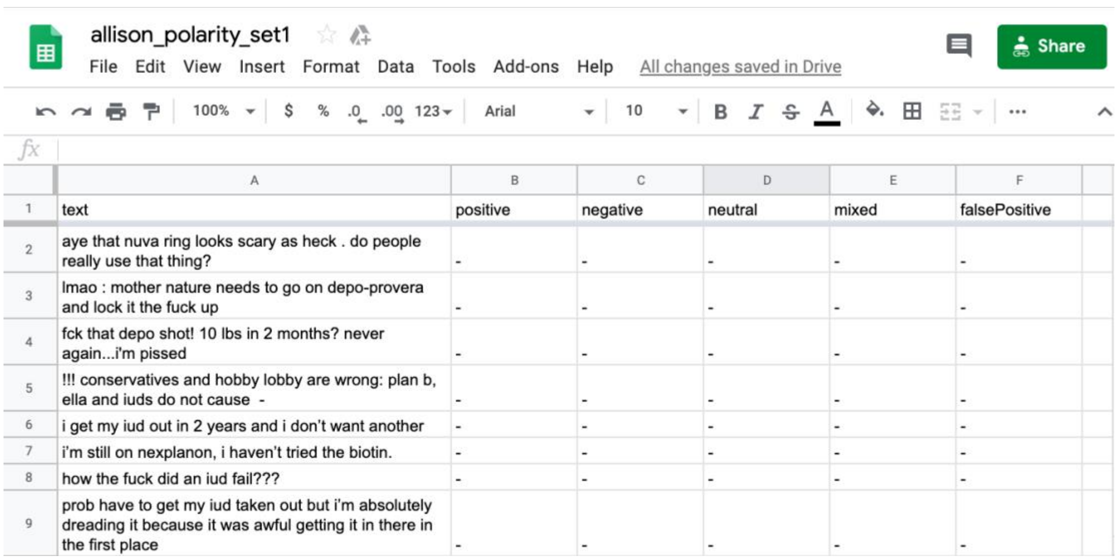
Ultimately, we'll compare the human analysis to the computer's in order to calculate the computer's sensitivity and specificity for a tweet's polarity with respect to the birth control method it mentions.

Working in Google Sheets

To actually analyze the tweets, you will work in google sheets. I will email each of you the link to a folder containing two google sheets that are unique to you. The first row will be filled in as an example, and highlighted in green. Before you start an "analysis session," browse through a chunk of tweets to get a 'lay of the land' prior to assigning sentiments - you may find that after reading a bunch of tweets, what you first thought was positive' feels more like 'mixed' after reading more. I will also have access to it so that we can download its contents for analysis once you've finished. It is VERY IMPORTANT that you work in the google sheet in order to avoid compatibility issues between different computer operating systems, versions of excel, numbers, etc. for when we ultimately consolidate all of the results to run statistical analyses in R.

For the polarity analysis, the blank google sheet will look like this:

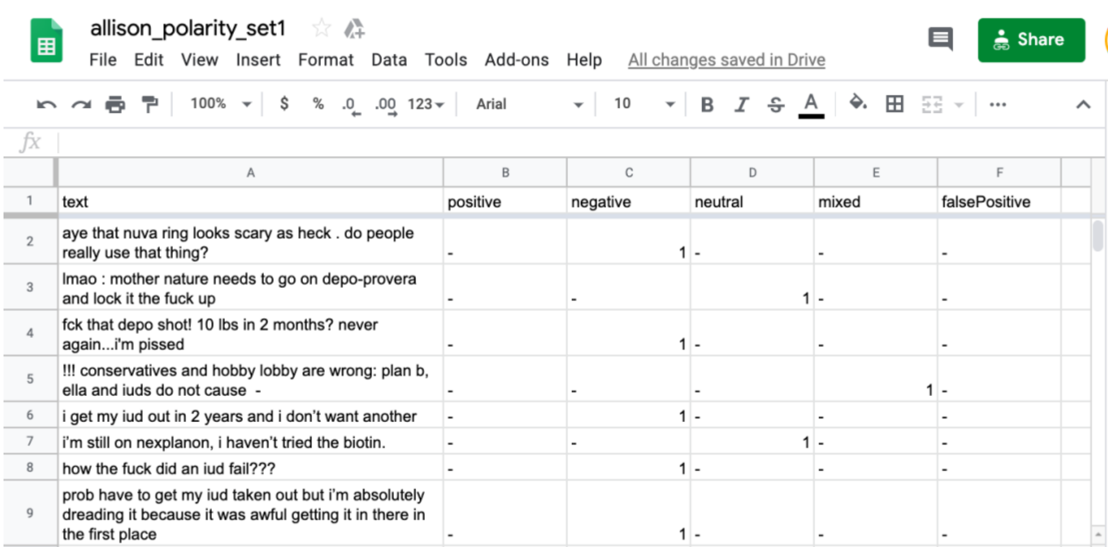
Figure 1: Blank polarity google sheet



	A	B	C	D	E	F
1	text	positive	negative	neutral	mixed	falsePositive
2	aye that nuva ring looks scary as heck . do people really use that thing?	-	-	-	-	-
3	lmao : mother nature needs to go on depo-provera and lock it the fuck up	-	-	-	-	-
4	fck that depo shot! 10 lbs in 2 months? never again...i'm pissed	-	-	-	-	-
5	!!! conservatives and hobby lobby are wrong: plan b, ella and iuds do not cause -	-	-	-	-	-
6	i get my iud out in 2 years and i don't want another	-	-	-	-	-
7	i'm still on nexplanon, i haven't tried the biotin.	-	-	-	-	-
8	how the fuck did an iud fail???	-	-	-	-	-
9	prob have to get my iud taken out but i'm absolutely dreading it because it was awful getting it in there in the first place	-	-	-	-	-

You will put a "1" in the appropriate column for each tweet, and leave the other cells with dashes (i.e. don't change them), like this:

Figure 2: Completed polarity google sheet



	A	B	C	D	E	F
1	text	positive	negative	neutral	mixed	falsePositive
2	aye that nuva ring looks scary as heck . do people really use that thing?	-	1	-	-	-
3	lmao : mother nature needs to go on depo-provera and lock it the fuck up	-	-	1	-	-
4	fck that depo shot! 10 lbs in 2 months? never again...i'm pissed	-	1	-	-	-
5	!!! conservatives and hobby lobby are wrong: plan b, ella and iuds do not cause -	-	-	-	1	-
6	i get my iud out in 2 years and i don't want another	-	1	-	-	-
7	i'm still on nexplanon, i haven't tried the biotin.	-	-	1	-	-
8	how the fuck did an iud fail???	-	1	-	-	-
9	prob have to get my iud taken out but i'm absolutely dreading it because it was awful getting it in there in the first place	-	1	-	-	-