Proceedings of the Eighteenth International AAAI Conference on Web and Social Media (ICWSM 2024)

Sensemaking about Contraceptive Methods across Online Platforms

LeAnn McDowall[1], Maria Antoniak[2], David Mimno[1]

[1] Cornell University
[2] Allen Institute for AI

Abstract

Selecting a birth control method is a complex healthcare decision. While birth control methods provide important benefits, they can also cause unpredictable side effects and be stigmatized, leading many people to seek additional information online, where they can privately find reviews, advice, hypotheses, and experiences of other birth control users. However, the relationships between their healthcare concerns, sensemaking activities, and online settings are not well understood. We gather texts about birth control shared on Twitter and Reddit—popular communities with different affordances, moderation, and audiences—to study where and how birth control is discussed online. Using a combination of topic modeling and hand annotation, we identify and characterize the dominant sensemaking practices across these platforms, and we create lexica to draw comparisons across birth control methods and side effects. We use these to measure variations from survey reports of side effect experiences, highlighting topics that social media users discuss more than expected online. Our findings characterize how online platforms are used to make sense of difficult healthcare choices, including analyzing risks, calculating timing and dosages, hypothesizing about causes of side effects, and storytelling about painful experiences. We contribute both to understanding unmet needs of birth control users and to exploring context-specific patterns in social media discussions.

Introduction

For many people, birth control is more than just contraception; it plays a critical role in health and well-being. In addition to providing a means of family planning, birth control methods can be used to treat and manage many medical conditions, including acne, endometriosis, cancer risks, gender dysphoria, and irregular and painful menstruation (Schindler 2013; Nahata, Chelvakumar, and Leibowitz 2017). However, birth control methods are not one-size fits all, and the choices of whether to use birth control and which method to select are complicated by personal beliefs, cost and accessibility, and a wide array of side effects that are difficult to predict and identify (Yoost 2014; Manzer and Bell 2022).

When navigating this decision, birth control users face a *sensemaking* challenge, a contextual process in which individuals work with a community to gather information, compare stories, and make sense of a shared experience (Weick, Sutcliffe, and Obstfeld 2005; Mamykina, Nakikj, and Elhadad 2015; Andalibi and Forte 2018; Young and Miller 2019). In the case of birth control, this challenge also involves dealing with social stigma, understanding contraceptive methods and potential side effects, and seeking normalcy after upsetting and painful experiences. This challenge leads many birth control users to the internet (Yee and Simon 2010; Russo et al. 2013), where they can engage in community sensemaking activities, e.g., seeking and sharing information, reviews, advice, hypotheses, and stories. While researchers have examined the frequencies of birth control discussions (Nobles, Dredze, and Ayers 2019; Merz et al. 2020), the sensemaking practices in these communities (and how these differ from other online healthcare communities) are unexplored.

Prior work has shown that these various sensemaking activities can vary across different online platforms (De Choudhury, Morris, and White 2014; Rivas et al. 2020; Zhang, N. Bazarova, and Reddy 2021), and social media traces are prone to context-dependent biases (Olteanu et al. 2018). Different healthcare conditions (Sannon et al. 2019) and side effects (De Choudhury, Morris, and White 2014) can result in different behavior, and recent work has demonstrated differences in the online sensemaking practices surrounding reproductive healthcare topics like pregnancy and vulvodynia (Young and Miller 2019; Andalibi and Garcia 2021; Chopra et al. 2021). But so far, birth control discussions have been examined only on individual platforms.

We study birth control discussants' unique combination of sensemaking activities—deciding between methods, interpreting side effects, calculating risks, etc.—and we use computational text analysis methods to discover the intersections between these activities and online platforms. Because different methods and side effects could lead to different sensemaking activities, we first explore *which birth control methods and side effects are more likely to be discussed on different online platforms* (**RQ1**). Using lexica crafted for our online settings, we compare these discussion frequencies to the distribution of experienced side effects as reported via a nationally representative survey by Nelson et al. (2017), indicating increased or decreased interest in discussing these concerns online. Next, using topic modeling and hand-annotations, we discover and characterize *the* *kinds of sensemaking activities that birth control discussants engage in online and how these differ by platform* (**RQ2**). We contrast these activities with prior work on related online healthcare communities, and we highlight the particular strategies of birth control discussants.

Contributions. We identify and characterize how birth control discussants make sense of their experiences and options via two large English-language online platforms: Reddit and Twitter.

- We identify a unique combination of sensemaking strategies including *storytelling*, *risk analysis*, *timing and calculations*, *causal reasoning*, *method and hormone comparison*, and *information and explanations*.
- Across platforms, we find that *storytelling* is used to prepare for and overcome painful insertion experiences.
- We find that Twitter users are more likely to discuss the IUD and severe side effects like stroke, while Reddit users frequently discuss both the IUD and the pill as well as sensitive side effects like bleeding.
- We compare our results to self-reported survey data, highlighting side effects that Twitter and Reddit users discuss more than expected, perhaps indicating increased interest and needs met by specific platforms.

Throughout this study, we employ a mixture of customized lexica, which we release to the public,[1] topic modeling, and hand annotation, demonstrating how a detailed and diverse analysis across online platforms can expand our understandings of social sensemaking, web and social media activity, and critical healthcare issues.

Related Work

Sensemaking in online communities. As explained by Weick, Sutcliffe, and Obstfeld (2005), sensemaking necessarily involves communication and is an "*activity that talks events and organizations into existence*," in part through retrospective storytelling. It is a process that relies on collaborative problem solving (Pirolli and Card 2005), and more recently, it has been defined in closely related work by Andalibi and Garcia (2021) as "*how individuals make sense of complex phenomena by constructing mental models that draw on new or existing experiences, information, emotions, ideas, and memories*." In online healthcare communities, the individual works to make sense of their healthcare experience, and the community collectively gathers information, compares stories, and makes sense of a shared experience (Mamykina, Nakikj, and Elhadad 2015). Sharing experiences can help narrators make sense of their stories (Tangherlini 2000) and can transfer important information to others without first-hand experience (Bietti, Tilston, and Bangerter 2019), leading users through a transformative process via the gathering and organizing of information (Genuis and Bronstein 2017; Patel et al. 2019). This process is contextual; each community can rely on different strategies (Young and Miller 2019).

Healthcare activity across platforms. Specific platform affordances can facilitate different types of healthcare-related disclosure and interactions. Zhang, N. Bazarova, and Reddy (2021) formulate a *social media disclosure ecology* in which platform affordances like anonymity, persistence, and visibility control can predict, e.g., pandemic-related disclosures. Health information seeking behavior can differ across platforms, with search engines more frequently used for serious and stigmatized conditions and Twitter more frequently used for symptoms (De Choudhury, Morris, and White 2014), and community affordances such as threaded conversations and hashtags can influence participation decisions of those with invisible chronic health conditions (Sannon et al. 2019). The distribution of content type can also differ by platform: e.g., prior work found that *sharing experiences* is less frequent on social networks than forums, although these rates can depend on healthcare topic (Rivas et al. 2020).

Reproductive healthcare and online sensemaking. Studies of reproductive healthcare communities have emphasized some unique sensemaking practices. For example, in a study of sensemaking practices after pregnancy loss, Andalibi and Garcia (2021) highlighted the importance of seeking emotional validation, rather than just information, when trying to re-discover normalcy. Young and Miller (2019) similarly supported these dual information management and emotional needs in a study of a vulvodynia Facebook group, and Chopra et al. (2021) discussed polycystic ovary syndrome and why personal tracking is an important communal activity that involves comparison to others' experiences. We investigate whether these sensemaking themes hold for online birth control communities.

Two recent studies have focused specifically on online discussions of birth control. First, Nobles, Dredze, and Ayers (2019) monitored Google search queries to track interest in different birth control methods, finding that U.S. political events correlate strongly with increased information seeking behavior online. Second, Merz et al. (2020) examined the prevalence and sentiment of tweets mentioning different birth control methods. The authors found that long-acting methods like the IUD were mentioned more often on Twitter, and the proportion of these tweets increased over time. While most tweets expressing sentiment about contraception were negative, tweets about long-acting methods were more likely to express positive sentiment. Other quantitative studies have similarly focused on sentiment (Pillarisetti et al. 2022; Álvarez-Mon et al. 2020) or frequency measurements of user engagement (Gurman and Clark 2016; Latack et al. 2021). To our knowledge, prior work studying online discussions of birth control has neither explored users' sensemaking practices nor compared patterns across methods, side effects, and platforms.

Data

We collected data related to birth control from two prominent online platforms: posts and comments on the Reddit community r/BirthControl and Twitter posts and replies mentioning birth control. Table 1 summarizes these datasets.

[1] https://github.com/lhmcd/birth-control-across-platforms

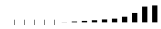
Community	# of Posts	Vocab Size	Mean Tokens	Year Range	Posts Dist. (2007-2020)	Moderation	Structure
Reddit Posts	68,958	49,088	79	2010-2020		user moderators	forum posts
Reddit Comments	264,912	67,837	32	2010-2020		user moderators	replies to forum posts
Twitter Posts	499,796	398,910	12	2006-2020		company	tweets (no retweets)
Twitter Replies	211,896	73,896	12	2007-2020		company	replies to tweets

Table 1: Overview of the two datasets, including only texts mentioning our target birth control methods (pill, IUD, implant).

Platform Selection

We focus our study on Twitter and Reddit, two large platforms that host public discussions about a wide range of topics, from politics to memes to healthcare. Discussions about birth control on these platforms include hundreds of thousands of users and over a decade of content (see Table 1).

While some subreddits and Twitter accounts specialize in visual content, both platforms are primarily text-based. This makes comparison feasible, while differences between the platforms (e.g., the ability to create topic-focused and user-moderated communities on Reddit, discoverability and wide audience on Twitter) make their comparison fruitful. Birth control discussions occur on many platforms, including Facebook, Instagram, YouTube, and TikTok, but research access to these platforms is limited and their frequent video content (a) is more difficult to automatically process and (b) allows for user interactions (e.g., reuse of trending sounds) distinct from text-based platforms.

Twitter and Reddit have been the frequent focus of prior research on online healthcare discussions, with Reddit studies examining reproductive healthcare topics like vulvodynia (Young and Miller 2019) and polycystic ovary syndrome (Chopra et al. 2021) and Twitter studies examining the frequencies of themes and sentiment of birth control discussions (Merz et al. 2020; Pillarisetti et al. 2022; Gurman and Clark 2016; Álvarez-Mon et al. 2020). Despite their popularity among both users and researchers, prior work has not directly compared birth control discussions across the two platforms, nor has prior work studied the sensemaking practices of birth control discussants on either platform.

Data Limitations

Our focus on Twitter and Reddit leads to some limitations. These are English-language subcommunities of larger websites, and while we have limited demographic information, we observe that the majority of location-specific posts (e.g., politics, insurance) are about the U.S. Where possible, we have included prior work describing general demographic observations on these platforms. Birth control can be used by many different people for a variety of healthcare reasons, including gender dysphoria, and we do not assume the gender of the birth control discussants.

Importantly, Twitter and r/BirthControl users included in this study might or might not use birth control; some of these users might be discussing birth control with no intent or history of seeking medication or interventions. Our goal is to compare discussions across platforms, not to predict offline rates of birth control usage or side effects.

Data Collection

Reddit. r/BirthControl[2] is a user-created and user-moderated online forum dedicated to birth control. Users are pseudonymous and range from one-time questioners to experienced question answerers. As of April 1, 2024, r/BirthControl had 136K members. According to a Pew Research Center survey of U.S. residents, more men (15%) than women (8%) use Reddit, and more white (12%) and Hispanic (14%) than Black (4%) survey respondents use Reddit (Perrin and Anderson 2019), but these platform-wide distributions are unlikely to be representative of r/BirthControl, for which more detailed demographics are unavailable. Prior work on a related subreddit found that 81% of the users identified themselves as white (Nobles et al. 2020).

Using the PushShift API, we collected the title (for posts), text, date, and user-assigned tag (for posts). We removed comments written by the parent post's author, and we also removed stickied comments (auto-generated or mod-written comments). We did not include user-deleted documents.

Twitter. Twitter is a large social network where discussions range widely from personal to global topics. Compared to the general public, Twitter users are more likely to be Democrats, and they also skew younger than users of YouTube, Facebook, or Instagram (Perrin and Anderson 2019). The gender and racial distribution on Twitter are close to uniform; out of a set of U.S. survey respondents, 24% of men and 21% women report using Twitter, while 24% of Black, 25% of Hispanic, and 21% of white respondents report using Twitter (Perrin and Anderson 2019). Birth control discussions take place in the context of many other discussion topics, and moderation is organized by Twitter.

Using the Twitter Academic API (v2), we collected all the English Twitter posts containing a set of keywords corresponding to our three target birth control methods. Separately, we collected all the Twitter replies containing the same set of keywords. Prior work has shown that this API returns reliable representations of the full tweet space (Pfeffer et al. 2022). We removed Twitter handles from tweet texts. The high vocabulary size for Twitter posts (Table 1) partly reflects many unique URLs shared in these documents, which can be indicative of information-providing behavior. There are many possible design decisions when gathering Twitter data, and we chose not to collect replies and

[2]https://www.reddit.com/r/birthcontrol/

parent posts to (a) limit the size of our collection, (b) target texts explicitly discussing birth control, and (c) more closely replicate prior work (Merz et al. 2020).

Methods

To answer RQ1, we developed lexica to track discussions of birth control methods and side effects. We decided on target methods and side effects via a collaborative process, beginning with existing frameworks and iteratively modifying as we examined our collected data. Because of our focus on sensemaking, our goal with these lexica was to measure *prevalence of discussions* rather than *prevalence of offline usage or experience* because (1) we are interested in the level of interest and concern of birth control discussants regardless of offline experiences and (2) other study designs (e.g., surveys) are better suited to studying usage rates. The frequency of discussions could be influenced by the level of *concern* birth control discussants on this platforms have about particular methods and side effects, biases introduced by the population or behavior unique to a platform setting (Olteanu et al. 2018), and a user's individual goals. We investigate these biases not only by comparing the platforms to one another but also by comparing their relative frequencies to self-reported side effect frequencies from survey studies.

To answer RQ2, we train and collaboratively code a pair of topic models to discover the sensemaking activities that are specific to birth control discussions on Reddit and Twitter. Lexica cannot capture the complex themes identified in prior work on non-birth control communities, such as storytelling and sharing/gathering information (Andalibi and Garcia 2021; Mamykina, Nakikj, and Elhadad 2015). Because we expect that prior sensemaking frameworks, which have not focused on birth control discussions, might not extend to this setting and the difficult healthcare choices faced by birth control users, we take a bottom-up approach, allowing the data-driven topics to guide our coding, as in prior work on qualitative analysis via topic modeling (Nelson 2020; Baumer et al. 2017).

Birth Control Method Lexicon

To explore the prevalence across platforms of different birth control methods, we develop a lexicon to match each document to the primary method it discusses, where "primary" is the method that is mentioned most frequently in the document.

Contraceptive Method Selection. Due to data scarcity and space limitations, we do not consider methods that are not available on both platforms and of the remaining methods, we limit our analysis to the three most commonly discussed reversible, non-emergency methods: the pill, the IUD and the implant, following prior work (Merz et al. 2020).[3]

Lexicon Development. We rely on sets of keywords to assign each text a primary birth control method. Because of their different structures, each platform requires its own set of keywords and matching techniques. All of these lexica are available in the Github repository linked above. To estimate performance, we manually checked 450 documents balanced across Reddit posts, Twitter posts, and Twitter replies.[4] Our precision and recall scores were perfect (1.0).

We initialize each lexicon with the medications listed as "pregnancy contraceptives" on WebMD, a healthcare review website. Links to these medications are listed on a centralized page on WebMD[5], and we extract the *drugname* parameter from each of the links. This parameter includes the medication name followed by the method type. For example, for the *drugname* "implanon-implant," the medication "implanon" would be added to our lexicon under the method "implant." We then augment or limit the lexica as noted below for each platform.

Reddit. To assign Reddit posts and comments to their respective birth control methods, we augment the lexicon derived from WebMD with terms from the Twitter keyword set and general terms like "pill" that are typically ambiguous on Twitter, but usable in the context of the *r/BirthControl* subreddit. We iterate on the keyword set, running our classifier and examining unassigned documents until we can no longer assign additional documents to a method. In all cases, the text was assigned to the method that had the highest keyword count (of all the methods).[6] If the highest keyword count was not for one of our target three methods, or if multiple methods were mentioned with equal frequency, we discarded the text. We follow a similar procedure for Reddit comments, but in cases where the comments do not contain a birth control keyword, we assign the comment to the birth control method of its parent post.

We were able to assign all but 4.6% of the Reddit posts and 32.6% of the Reddit comments to a birth control method. Manual examination of the unassigned posts reveals discussions of pregnancy scares, access (e.g., online appointments and prescriptions), and treatment of side effects with non-contraceptive medications. The remainder of these unassigned posts discussed topics that are adjacent to birth control, e.g., insurance, menstruation, but are not explicitly connected to a birth control method.

Twitter. We use a more limited keyword set, derived from a similar survey of birth control tweets by Merz et al. (2020), to query for tweets about each of our target birth control methods. Our focus was on precision rather than recall, as the Twitter API requires a keyword match to retrieve tweets

[3]Using the lexica described above for Reddit, we find that the implant (8,578 posts) is discussed more than twice as often as the shot (3,529 posts) or barrier methods like condoms (3,029 posts). This contrasts with reported rates of contraception use in the U.S., where the pill is used by 14.0%, the IUD by 8.4%, condoms by 8.4%, the implant by 2.0%, and the shot by 2.0% of women aged 15-49 (Daniels and Abma 2020).

[4]We omitted Reddit comments because of our method detection strategy, which defaulted to the method in the parent post if no method was found in the comment.

[5]https://www.webmd.com/drugs/2/condition-3454/pregnancy-contraception

[6]We could instead have assigned each Reddit post and comment to as many methods as were mentioned in its text, rather than assigning each post and comment only to the method mentioned most frequently. We do not find that these different assignment techniques affect our distributional results.

Source	# Side Effect Texts	% Side Effect Texts
Reddit Posts	53,027 / 72,731	73%
Reddit Comments	119,780 / 238,568	50%
Twitter Posts	61,698 / 513,017	12%
Twitter Replies	47,049 / 244,140	19%

Table 2: The coverage of the side effects lexicon.

on a particular topic; if we included keywords like "pill", our results would contain many false positives, unlike Reddit where the topic is already constrained to birth control. After gathering our initial Twitter dataset, we then apply the full Reddit keyword set using the same methods described for Reddit above. As with the Reddit data, we find that more texts can be assigned to only the pill, IUD, or implant than to a combination of methods; 12.7% of posts and 17.6% of the gathered tweets either mentioned multiple methods with equal frequency or most frequently mentioned a method not in our set of three target methods.

Side Effects Lexicon

To measure the frequency of discussions about birth control side effects, we develop a lexicon of terms and patterns. Because we do not use the side effects lexicon for data collection (unlike the methods lexicon), we rely on one lexicon across both datasets. We select patterns that match any discussion of the side effect, whether or not it is affirmative.

We grounded our development of the side effects lexicon in prior work. We matched the side effect categories from Nelson et al. (2017) as closely as possible; this study conducted a nationally representative survey of U.S. birth control experiences and reported prevalences of side effect experiences across different birth control methods. In addition to these, we added lexicon categories for pain, skin changes, PMS, appetite changes, sexual partner feeling IUD strings, and heart attack. We identified these additional categories and patterns from topic models trained on our datasets and from other work; for example, we also drew on work by Barr (2010) that discusses a variety of side effects and their known frequencies and associations with different birth control methods. For each side effect, we then iteratively queried and made updates to the lexicon when we encountered false positives or false negatives.

Lexicon evaluation. To evaluate our lexicon, for each side effect we randomly selected a set of matching and non-matching texts, balanced across platforms and match status, resulting in a set of 1,040 texts. We then manually checked whether each text does or does not contain a discussion of the target side effect. Across the side effects, we found precision of 0.98 and recall of 0.96. We show the lexicon coverage in Table 2.

Comparing observations with survey responses. We compare the distributions of side effect mentions to prior work surveying people in the U.S. about their experiences with birth control side effects. This allows us to compare the frequency of side effect experiences with the frequency of online discussions. Differences between these distributions can indicate healthcare needs that users are addressing via the internet. We compare to surveys (rather than controlled studies) because surveys better approximate the self-reported anecdotes and personal experiences shared online. We first converted each distribution to a ranking, where lower ranks represent greater percents of discussions $r_{platforms}$ or experiences r_{survey}. We then find the difference between the ranks for each side effect, d_{ranks}.

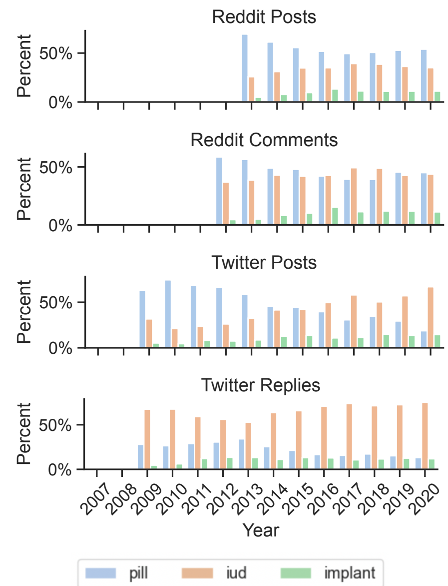


Figure 1: Document distributions of methods over time. We show only years with at least 1000 posts or comments for at least one method.

$$d_{ranks} = r_{survey} - r_{platforms} \quad (1)$$

These ranks avoid issues in directly comparing percents, which might on average be higher or lower. When d_{ranks} is positive, it indicates more online discussion than expected given the frequency of reported side effect experiences.

Sensemaking Topic Model

We compare sensemaking activity distributions through a bottom-up analysis. While some sensemaking themes might be shared across settings, other themes might be more prevalent in certain platforms and side effects. We use an unsupervised, automatic method to help us avoid biases in our data coding (which might overlook certain themes while over-representing others) by quickly revealing frequent themes across all the texts in each dataset (Nelson 2020).

For this analysis, we rely on topic modeling, an automatic method that identifies prominent themes and discourses in a dataset (Blei, Ng, and Jordan 2003). This unsupervised model can provide a quick and data-driven way to explore a dataset (Chang et al. 2009). Unlike purely qualitative methods, topic modeling allows for analysis of large pools of data (Baumer et al. 2017), and unlike supervised methods,

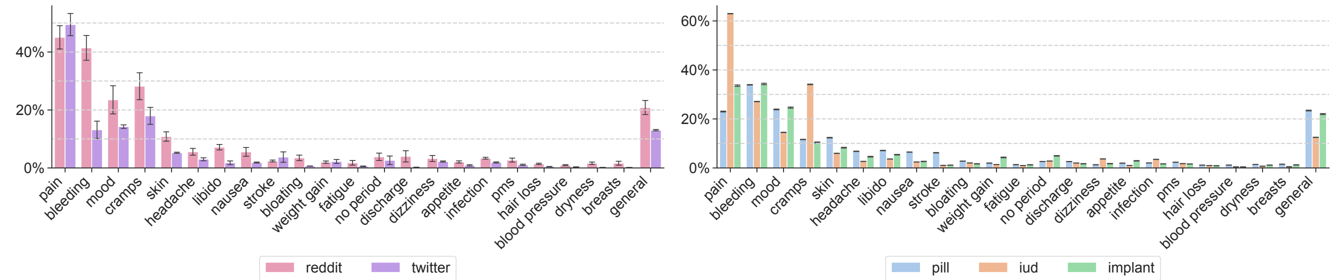


Figure 2: Distribution of the side effect mentions across the platforms and methods (after 2012). Bars represent the percents of documents for the specified platform or method mentioning *any* side effect that also mention the *specified* side effect. See Table 2 for the denominator sizes. Error bars indicate the standard deviation over 20 bootstrapped samples of the datasets.

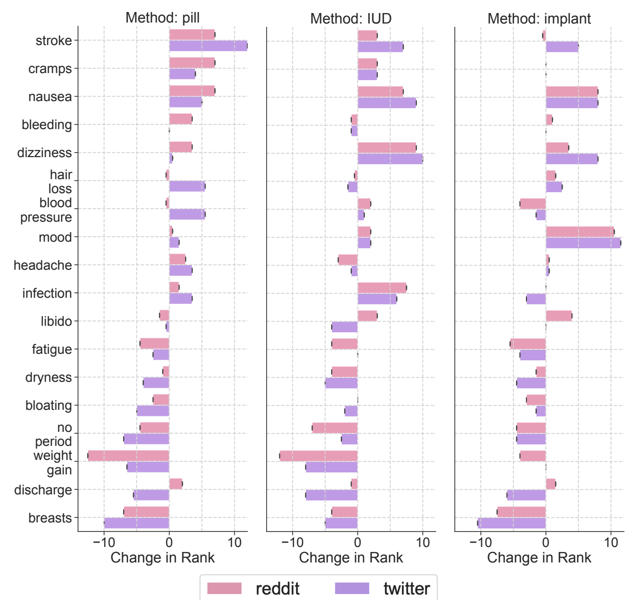


Figure 3: Distribution of the side effect mentions across the platforms. Bars represent the *differences* between the rank reported in the survey results from Nelson et al. (2017) and the rank on the specified platform. Platform ranks are determined by first finding the frequency of side effect mentions; these are the percent of documents mentioning *any* side effect that also mention the *specified* side effect. Bars to the *right* of the x-axis represent side effects that are mentioned *more frequently* on the platforms than are reported in the survey. Results are shown only for texts posted after 2012, and error bars indicate the standard deviation over 20 bootstrapped samples of the datasets.

topic modeling supports a bottom-up analysis that combines human interpretation (via manual evaluation and labeling) with automatically-recognized patterns (Nelson 2020). Topic modeling continues to be a popular technique to analyze online health communities (Nobles et al. 2018; Yang et al. 2019; Abebe et al. 2020; Nobles et al. 2020).

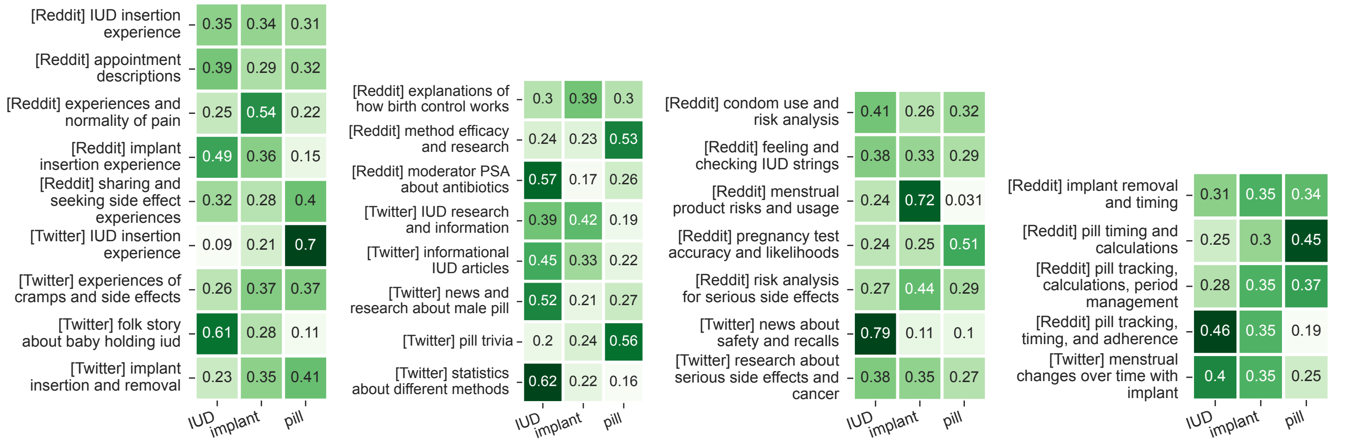
Model training. We trained a latent Dirichlet allocation (LDA) topic model (Blei, Ng, and Jordan 2003) on each of our datasets, by combining the Reddit posts and comments and the Twitter posts and replies for training. We use MALLET[7] for training. LDA remains a highly consistent and reliable model (Harrando, Lisena, and Troncy 2021; Hoyle et al. 2022), especially when trained via Gibbs sampling for smaller datasets. We balanced each training set, sampling 8,000 documents for each method for each of the Reddit and Twitter datasets (see Figure 1 for method distributions). By balancing the training set, we avoid weighting the topics toward a certain method.

We removed a set of frequent stopwords, numbers, punctuation, and duplicate documents from the training sets. Removing stopwords and duplicate documents has been shown to improve the legibility of the final topics, whereas stemming can be harmful (Schofield and Mimno 2016; Schofield, Magnusson, and Mimno 2017; Schofield, Thompson, and Mimno 2017). To avoid capitalization discrepancies, we lowercase all text. We experimented with different numbers of topics and found 35 to be interpretable across the datasets; at this number of topics, we observed topics that were neither too fine-grained (e.g., splitting a single theme across multiple topics) nor too high level (e.g., combining themes that should be separated), and that produced reasonable evaluation scores (see below). However, we emphasize that there is no “correct” number of topics and that this method is used for exploration and interpretation.

Model evaluation. While we do not have space here to list the full sets of topics for each dataset, we will provide the topics, most probable words, and example paraphrased documents in our code repository for manual examination. We report human evaluation of our topics following the recommendations in Hoyle et al. (2021). Using the *word intrusion task* (Chang et al. 2009), we show one non-expert annotatorand two expert annotators (the first two authors) a set of the four most probable words plus an “intruder” word that has low probability for the current topic but high probability for another topic. We report the proportion of topics for which the annotators identified the intruder.

We found that our Reddit topics have performance much higher than a random baseline of 0.2 (annotator accuracies for Reddit: 0.71, 0.74, 0.77) while the Twitter topics

[7]http://mallet.cs.umass.edu/



(a) Storytelling (b) Information & Explanations (c) Risk Analysis (d) Timing & Calculations

Figure 4: Users engage in different sensemaking strategies when discussing different birth control methods. For example, users engage in more storytelling about the IUD and implant than the pill. Cells show the mean topic probabilities across methods for four of our sensemaking clusters. The color scale is constant across the plots, and darker colors indicate higher probabilities; these topics and methods are more likely to co-occur. To highlight differences, rows are normalized to sum to one.

have lower performance but are still substantially above the random baseline (Twitter: 0.46, 0.46, 0.51). The first (non-expert) annotator consistently had lower scores. The lower performance on Twitter is expected, as text processing methods are notoriously challenged by the short text lengths and non-standard language (Gimpel et al. 2011), and the short tweets require contextual knowledge to interpret. This vulnerability of the word intrusion task to esoteric topics is noted by Hoyle et al. (2021).

We also calculate the "UMass" Coherence: the log probability that a document containing at least one instance of a higher-ranked word also contains at least one instance of a lower-ranked word (Mimno et al. 2011; Röder, Both, and Hinneburg 2015). We find the higher mean scores across topics for Reddit (-416) and lower scores for Twitter (-712). We note the criticisms of coherence and automatic metrics Hoyle et al. (2021); in comparison to human evaluation, automatic metrics can exaggerate model differences.

Identifying sensemaking themes. Across the datasets, we find that many of the topics discuss birth control methods (including one method in their 10 most probable words), side effects, pregnancy, and access (costs, appointments). We then identify a series of cross-cutting sensemaking topics. These topics include discussions of information seeking/sharing, educational links and resources, how-to explanations, experience seeking/sharing, emotional support, and other sensemaking-related discussions. Two researchers independently coded the topics as more or less related to sensemaking. The researchers then conferred and agreed upon a final set of topics and assigned them to thematic clusters. During this coding, we relied on the sensemaking definition from Andalibi and Garcia (2021) provided above. We also drew inspiration from prior work studying sensemaking in online healthcare communities, particularly those works also focused on reproductive healthcare (see §Related Work). We further explore and validate these sensemaking topics by hand-labeling a small subsection of the data with social support goals (Yang et al. 2019). After coding 150 documents for each dataset, we measured the agreement between the annotators using Krippendorff α, as each document could receive zero, one, or more labels. Our agreement was acceptable, with a score across the labels of 0.74.[8]

To compare the topics across the platforms, we aligned the topics from the different models using Jensen-Shannon divergence (JSD) for the word distributions associated with each pair of topics. After manual examination of the ranked topic pairs, we categorized topics with JSD scores below 0.6 as aligned across the datasets, and those with scores above 0.8 we considered diverging.

Results

Results: Birth Control Methods Across Platforms

Figure 1 shows the distributions of the methods across the different platforms, where frequency proportions are calculated by dividing the number of texts mentioning the specified side effect by the total number of texts mentioning any side effect (i.e., $p(s_i|a)$ where s_i is the specific side effect and a is any side effect). Discussions of different methods have changed dramatically over time. On Twitter, the pill begins as the most popular but is replaced by the IUD in posts, while replies always center on the IUD. Twitter replies also increase sharply after 2016, differing from the Twitter posts (so is unlikely to be a symptom of our keywords). This suggests that the IUD generates more discussion on Twitter, especially post-2016, compared to the other birth control

[8]Lower scores are not surprising for subjective language labeling tasks (Artstein and Poesio 2008; Godwin and Piwek 2016), and our scores are substantially higher than the agreement scores for a very similar classification task in Rivas et al. (2020).

methods. On Reddit, the number of posts discussing the pill and IUD are similar, with slightly more posts discussing the pill, though this difference is erased in the comments. The implant is always the least discussed of the three methods.

Results: Side Effects Across Platforms

Figure 2 shows the distribution of side effects across platforms. We find that pain, cramps, menstrual bleeding irregularities, mood changes, and skin conditions are the most commonly discussed side effects on both platforms. In particular, *pain is consistently and frequently discussed across the platforms* and is an outlier among the side effects. Discussions of stroke are more frequent in Twitter posts than in the other datasets, but for all other side effects, Twitter has the lowest discussion frequencies (perhaps because of stigma around publicly sharing such information compared to the smaller community and pseudonymous settings on Reddit). In contrast, Reddit has a much higher frequency of discussions for menstrual bleeding irregularities, mood changes, and general discussion of side effects (i.e., mentioning the term *side effect*).

Figure 2 also compares the side effects by birth control method. Across the birth control methods, we again find that pain is a frequently mentioned side effect, but it is most frequently mentioned in discussions that also mention the IUD. The implant is the only method shown to cause weight gain (Barr 2010), so the discussions of weight gain for the other methods are less expected and could indicate that this potential side effect is a concern across methods. Menstrual bleeding, mood changes, headache, and libido are least often discussed with the IUD, while nausea, stroke, and skin conditions are most often discussed with the pill.

Results: Comparison to Observed Distributions

Figure 3 shows the differences between our observed distributions of side effect mentions and the reported distributions of side effect experiences in Nelson et al. (2017) (described above). We find large differences between how frequently side effects are discussed online compared with how frequently they are reported in the survey from Nelson et al. (2017). For example, strokes are rarely experienced according to Nelson et al. (2017), but when mentioned with the pill, they are discussed more frequently on Twitter in comparison to other side effects, perhaps because sensational topics are frequently discussed on Twitter. Mood changes are more likely to be discussed across the platforms for the implant in comparison to the survey data. Dizziness is more likely to be discussed for the IUD than expected, while weight gain is universally discussed less frequently than expected. These patterns indicate cases in which users turn to the internet more or less frequently than expected based on their self-reported experiences in the survey.

Results: Sensemaking Across Platforms

Our final set of sensemaking themes included: **storytelling** (e.g., *implant insertion and removal* on Twitter), **risk analysis** (e.g., *news about safety and recalls* on Twitter), **timing and calculations** (e.g., *pill timing and calculations* on Reddit), **method and dose comparison** (e.g., *hormone dosages and comparison* on Reddit), **causal reasoning** (e.g., *weight changes and causes* on Reddit), and **information and explanations** (e.g., *research on the male pill* on Twitter). We show the four most frequent of these themes in Figure 4, with the topics from each platform included in that theme and their probabilities for each birth control method. We observe a greater number of storytelling topics for the IUD and implant than for the pill, and examination of these topics shows that insertion and removal experiences fuel this pattern. Each method is associated with risks and is included in method comparisons, but the pill in particular is included in discussions of timing and calculations.

Using Jensen-Shannon divergence to compare topics across the platforms, we find that the most **aligned** ($JSD < 0.6$) topic categories were: *weight changes*, *general side effects*, *IUD insertion experiences*, *implant insertion experiences*, *menstrual timing and cycles*, and *bleeding changes*. Each of our datasets included at least one representative topic from these categories.

The most **diverging** ($JSD > 0.8$) topics were: *causes and side effects of vaginal infections* and *explanations of how birth control works* on Reddit; and *IUD jokes and random*, *viral folk stories*, and *implant news about unexpected experiences* on Twitter. These topics were the most unique to their training dataset, without directly comparable topics in the other dataset.

Discussion

Our results indicate large differences in sensemaking strategies and discussions of methods and side effects across Twitter and r/BirthControl. Importantly, these patterns do not necessarily indicate real-world increases or decreases in use of different methods or experiences of side effects. The discrepancies we find between reported side effect experiences and rates of online discussions highlights the importance of studying online patterns as context-specific behavior. These patterns should not be used to forecast offline behavior but rather allow us to learn about platform-specific interests and strategies when discussing difficult healthcare topics.

Methods and side effects across platforms. In response to our first research question, we find that birth control discussions on Twitter and r/BirthControl substantially differ in their distributions of methods and side effect mentions. We cannot claim that using a method or experiencing a side effect *cause* people to choose a specific platform, but our observations add detail and sometimes contradict prior findings. For example, in a study of general healthcare information seeking, De Choudhury, Morris, and White (2014) found that people more often use search engines for serious and stigmatized conditions and more often use Twitter to discuss symptoms. We do not include search engines in our study, but in our comparison across platforms, we find that Twitter has a higher frequency of severe side effect discussions for birth control, while r/BirthControl has higher frequencies of general side effect discussions. Our focus on birth control might explain these variations, as many of the birth control side effects are themselves highly stigmatized (e.g., menstrual bleeding, vaginal discharge). The discussion of severe side effects on Twitter is likely related to Twitter's tendency toward sensational content, and could also explain the relative frequency of IUD discussion on Twitter; the IUD has been associated with both severe side effects and potential legal bans (Nobles, Dredze, and Ayers 2019). These findings emphasize the importance of considering platform setting when studying online patterns related to specific health conditions, medications, and side effects.

Online sensemaking practices of birth control discussants. In response to our second research question, our identification and comparison of topics that align with sensemaking themes reveals important activities specific to birth control and to the different platforms and methods. We identified seven categories of sensemaking activities, ranging from causal reasoning (especially about side effects) to comparisons of methods and hormone doses. Storytelling about the IUD and implant was a common practice, with strongly aligned pairs of topics across the platforms. But we find notable differences in the prevalence of *information & explanations* (more sharing of articles on Twitter), *risk analysis* (r/BirthControl includes instructions for how to assess risk, while Twitter highlights research and recalls), and other sensemaking practices. Overall, r/BirthControl is more suitable for inquiry into the more intimate nuances of birth control usage (e.g. checking IUD strings), while Twitter suits inquiry into information sharing practices (e.g. sharing news articles and research, announcing recalls). This combination of themes reflects the unique dilemmas facing birth control discussants, as they choose between "least bad" options, struggle to identify and treat side effects, determine risk of pregnancy given their circumstances, and avoid rare but alarming outcomes like stroke and heart attack.

Prior work has found that online communities focused on reproductive healthcare (e.g., pregnancy, vulvodynia) employ strategies related to validation (Andalibi and Garcia 2021), information management (Patel et al. 2019; Young and Miller 2019; Andalibi and Garcia 2021), personal tracking (Chopra et al. 2021), and identifying causation (Patel et al. 2019). We find these themes again in our birth control communities, with variations; for example, personal tracking is also a prominent theme in birth control discussions, but it is focused on pill timings and calculations and on self-observations of side effects. Identifying causation is a common concern for those experiencing infertility (Patel et al. 2019), and we find this theme again in our datasets but in the context of side effects like weight gain.

Storytelling, pain, and the IUD. We would like to particularly highlight storytelling as an important sensemaking strategy for birth control discussants. Storytelling is known to help communities work through trauma (Tangherlini 2000), and birth control discussants employ storytelling specifically to address physical pain. Pain is a major cross-cutting theme across the platforms; it is the most frequently and consistently discussed side effect, and it is most often mentioned alongside the IUD. While Barr (2010) identifies pain as the second most common side effect (after bleeding changes) and Dickerson et al. (2013) identifies pain as the most common side effect for the IUD and second most common side effect for the implant, we find a much larger gap between pain and the next most frequent side effects in our analysis. Pain is mentioned frequently in posts about the IUD and in posts whose probable topics are about seeking and sharing IUD insertion stories. Pain is inherently a subjective experience that cannot be precisely communicated (Scarry 1987), but sharing of personal stories provides one way for a community to build a sense of what is *normal* or *to be expected* (Patel et al. 2019; Andalibi and Garcia 2021). Our results suggest that there is an urgent unmet need for (a) honest education and preparation before IUD insertions and (b) pain treatment options during this procedure.

Online discussions differ from survey reports. In comparison to the survey results in Nelson et al. (2017), we find not only large differences in the frequency at which different side effects are discussed but also differences across the platforms. On one hand, these results demonstrate the risks in relying on social media traces to predict offline behavior (Olteanu et al. 2018). On the other hand, this comparison usefully highlights side effects for which interest is higher than we would have predicted using the survey data, indicating the contexts in which these discussants use social media to make sense of birth control.

It could be that the differences from the survey are due to the demographic distribution opting into the survey versus those opting to post online; future work surveying the demographic distribution of these users would address this question, but our results take a first step at measuring differences between these settings. We find, e,g., while mood changes are discussed at similar frequencies in comparison to the survey data across the platforms, other side effects like strokes, bloating, fatigue, bleeding, and dizziness might be discussed more or less frequently compared to the reports in Nelson et al. (2017), depending on the platform. Bloating is less frequently mentioned on all the platforms and for all the methods in comparison to the survey data, perhaps indicating a general lack of concern about this side effect in contrast to its reported frequency. But bloating is much less frequent on Twitter for the pill and implant, which could also indicate that Twitter is less suited to its discussion, perhaps because of embarrassment in the public setting or perhaps because of the more knowledgeable and helpful audience on Reddit.

Stigma, privacy, and contextual disclosures. Birth control can be a controversial, stigmatized, and intimate topic. This can lead birth control discussants to seek out additional information privately. For example, in a set of interviews of young Black and Hispanic women, Yee and Simon (2010) found that a greater number reported seeking decision-making support on the internet, citing its privacy, in comparison to other sources of information (e.g., talking to physicians, reading provided information). Interpreting side effects, analyzing the risk of pregnancy, or normalizing a painful experience require disclosing personal details and stories. Birth control discussants might analyze the risk and benefit of making these disclosures in certain settings and to certain audiences. While social penetration theory (Altman and Taylor 1973) posits that more disclosures are possible as social bonds deepen, prior work has also found that intimate language can be frequent among both close connections and strangers (but not in between) (Pei and Jurgens 2020). The variations we observe across methods, side effects, and sensemaking practices could be indicative of platform affordances for privacy, audience size, and anonymity, each of which can affect decisions to self-disclose. For example, we found that side effect discussions are less frequent on Twitter, where users are facing a much larger and non-specialized audience, unlike the other platforms. Giving users more platform-specific tools to control their audience size and membership (Mondal et al. 2014) could allow for more productive discussions of this sensitive topic.

Other factors influencing platform decisions. Research relying on social media data is vulnerable to various biases. For example, political heterogeneity of social media users can lead to imprecise models (Alkiek, Zhang, and Jurgens 2022), and user rating and connection behavior can shift after the introduction of new platform features (Malik and Pfeffer 2016). These risks can also vary based on research goals; Olteanu et al. (2018) draw a distinction between research that uses social data to study phenomena *beyond* social platforms (e.g., using our lexicon frequencies to predict offline popularity of these methods) versus research that uses social data to study phenomena *specific* to social platforms (e.g., using our lexicon frequencies to examine the sensemaking practices of Twitter or Reddit users). Our comparison to self-reported survey data highlights the limits of the first kind of research—patterns in online discussions of birth control differ significantly from offline patterns of experiences—while emphasizing the possibilities of the second research goal—such discrepancies help us characterize the discussions happening in these *specific* communities.

Decisions to seek information online can also correlate with demographic characteristics. For example, in a survey of U.S. young adults, those with a sexual risk history (early sexual activity, involvement in an unintended pregnancy) less frequently reported using the internet as a source and more frequently reported seeking information from a doctor/nurse, and men more frequently reported using the internet than women (Khurana and Bleakley 2015). It is also possible that users follow a *birth control journey*, where different needs at different points in one's journey can lead one to different platforms, as has been reported for other healthcare topics (Sannon et al. 2019; Andalibi and Forte 2018). These journeys can intersect with methods; for example, while the pill is a popular first method, many people report switching to the IUD as they gain more experience with birth control (Nelson et al. 2017). This would mirror the journeys of those with invisible chronic illnesses who move from one platform to another as their needs evolve and as they grow more comfortable with self-disclosure (Sannon et al. 2019). This is mirrored in intra-community research that models user trajectories in online cancer support groups, finding that users often transition from information-seeking to information-sharing roles over time (Yang et al. 2019).

Recommendations. As in past work exploring biases in social media data (Olteanu et al. 2018), our results show that birth control discussion patterns are often not consistent across platforms. For social media researchers interested in birth control discussions, our lexica can be used for Twitter and r/BirthControl but must be modified for new platforms and updated as new medications become available. When selecting a platform for birth control research, its choice should be guided by a matching between research goals and the method, side effect, and sensemaking distributions revealed in our results. More generally, social media research can follow our practice of mixing computational tools like topic modeling with lexicon-based methods and hand-annotation, using these methods to characterize difficult-to-identify themes like sensemaking practices. We recommend taking this careful approach, putting unsupervised results in context with qualitative and fine-grained measurements.

Broader Impacts, Ethics, and Limitations

Our study was considered exempt under Cornell University's IRB. However, while Reddit and Twitter posts and replies are "public," they can contain highly personal information, requiring a balance between potential harms and potential benefits to the community, as described in the guiding principles of the Belmont Report: *respect for persons*, *beneficence*, and *justice*.[9] Considering possible harms, e.g., re-identification of those using stigmatized medications, we do not collect any user-specific information, and we do not infer medical conditions for individual users; instead, we rely on patterns averaged across many users. We also anonymize and paraphrase any direct quotations. To protect users' agency to edit or delete their data at its original source, we release our data collection lexica but not copies of the collected data. We balance these concerns and protective actions against the benefits of this research; among other benefits, our work highlights the unmet pain treatment needs of a population known to be discriminated against by physicians (Samulowitz et al. 2018) and examines the kinds of support needed by those facing difficult healthcare choices. Finally, we do not assume or attempt to infer the gender of our dataset authors, as both birth control users and birth control discussants can include a diverse group of people.

Limitations

Social media does not necessarily represent offline events, and unlike the surveys we use for comparison, we cannot control for demographics. We study texts written in English and platforms that attract a U.S. audience, as the authors are all most familiar with this setting. We focus on one specific community out of many, and we do not expect that the findings in this paper will necessarily generalize outside of the U.S. or to other online spaces; indeed, our results indicate that different platforms display different patterns. Finally, averaging over posts allows us to track patterns and make comparisons but can also reduce nuance. Our work is best read in conjunction with ethnographic studies like Homewood and Heyer (2017) and Daley (2014), which highlight individual voices of birth control users.

[9]https://www.hhs.gov/ohrp/regulations-and-policy

Acknowledgements

We thank our anonymous reviewers for their helpful comments, and we also thank Sharifa Sultana, Katherine Antoniak, and other researchers for their feedback on this paper.

References

Abebe, R.; Giorgi, S.; Tedijanto, A.; Buffone, A.; and Schwartz, H. A. A. 2020. Quantifying Community Characteristics of Maternal Mortality Using Social Media. In *Proceedings of WWW*.

Alkiek, K.; Zhang, B.; and Jurgens, D. 2022. Classification without (Proper) Representation: Political Heterogeneity in Social Media and Its Implications for Classification and Behavioral Analysis. In *Findings of the ACL*.

Altman, I.; and Taylor, D. A. 1973. *Social penetration: The development of interpersonal relationships.*

Álvarez-Mon, M. A.; Donat-Vargas, C.; Llavero-Valero, M.; Gea, A.; Álvarez-Mon, M.; Martínez-González, M. A.; and Burgo, C. L.-D. 2020. Analysis of Media Outlets on Women´s Health: Interest on Contraceptives and their Side Effects, with Negative Perceptions of Tubal Ligation (Preprint).

Andalibi, N.; and Forte, A. 2018. Announcing Pregnancy Loss on Facebook: A Decision-Making Framework for Stigmatized Disclosures on Identified Social Network Sites. In *Proceedings of CHI*.

Andalibi, N.; and Garcia, P. 2021. Sensemaking and Coping After Pregnancy Loss: The Seeking and Disruption of Emotional Validation Online. In *Proceedings of CSCW*.

Artstein, R.; and Poesio, M. 2008. Inter-coder agreement for computational linguistics. *Computational Linguistics*.

Barr, N. G. 2010. Managing adverse effects of hormonal contraceptives. *American Family Physician*.

Baumer, E. P.; Mimno, D.; Guha, S.; Quan, E.; and Gay, G. 2017. Comparing grounded theory and topic modeling: Extreme divergence or unlikely convergence? *Journal of the Association for Information Science and Technology*.

Bietti, L. M.; Tilston, O.; and Bangerter, A. 2019. Storytelling as Adaptive Collective Sensemaking. *Topics in Cognitive Science*.

Blei, D. M.; Ng, A. Y.; and Jordan, M. I. 2003. Latent Dirichlet allocation. *JMLR*.

Chang, J.; Gerrish, S.; Wang, C.; Boyd-Graber, J. L.; and Blei, D. M. 2009. Reading tea leaves: How humans interpret topic models. In *Advances in NeurIPS*.

Chopra, S.; Zehrung, R.; Shanmugam, T. A.; and Choe, E. K. 2021. Living with Uncertainty and Stigma: Self-Experimentation and Support-Seeking around Polycystic Ovary Syndrome. In *Proceedings of CHI*.

Daley, A. M. 2014. What influences adolescents' contraceptive decision-making? A meta-ethnography. *Journal of Pediatric Nursing*.

Daniels, K.; and Abma, J. C. 2020. Current Contraceptive Status Among Women Aged 15–49: United States, 2017–2019. *NCHS Data Brief No. 388*.

De Choudhury, M.; Morris, M. R.; and White, R. W. 2014. Seeking and Sharing Health Information Online: Comparing Search Engines and Social Media. In *Proceedings of CHI*.

Dickerson, L. M.; Diaz, V. A.; Jordan, J.; Chirina, S.; Goddard, J. A.; Carr, K. B.; and Carek, P. J. 2013. Satisfaction, early removal, and side effects associated with long-acting reversible contraception. *Family Medicine*.

Genuis, S. K.; and Bronstein, J. 2017. Looking for "normal": Sense making in the context of health disruption. *JASIST*.

Gimpel, K.; Schneider, N.; O'Connor, B.; Das, D.; Mills, D.; Eisenstein, J.; Heilman, M.; Yogatama, D.; Flanigan, J.; and Smith, N. A. 2011. Part-of-Speech Tagging for Twitter: Annotation, Features, and Experiments. In *Proceedings of ACL*.

Godwin, K.; and Piwek, P. 2016. Collecting reliable human judgements on machine-generated language: The case of the QG-STEC data. In *Proceedings of INGL*.

Gurman, T.; and Clark, T. L. 2016. #ec: Findings and implications from a quantitative content analysis of tweets about emergency contraception. *Digital health*, 2.

Harrando, I.; Lisena, P.; and Troncy, R. 2021. Apples to Apples: A Systematic Evaluation of Topic Models. In *Proceedings of RANLP*.

Homewood, S.; and Heyer, C. 2017. Turned On / Turned Off: Speculating on the Microchip-Based Contraceptive Implant. In *Proceedings of DIS*.

Hoyle, A.; Goel, P.; Hian-Cheong, A.; Peskov, D.; Boyd-Graber, J.; and Resnik, P. 2021. Is Automated Topic Model Evaluation Broken? The Incoherence of Coherence. *Advances in NeurIPS*.

Hoyle, A. M.; Goel, P.; Sarkar, R.; and Resnik, P. 2022. Are Neural Topic Models Broken? *ArXiv*, abs/2210.16162.

Khurana, A.; and Bleakley, A. 2015. Young adults' sources of contraceptive information: variations based on demographic characteristics and sexual risk behaviors. *Contraception*.

Latack, K. R.; Yuen, F.; Wang, C.; and Nguyen, B. T. 2021. Online community queries on hormonal male contraception: an analysis of the Reddit "Ask Me Anything" experience. *Contraception*.

Malik, M. M.; and Pfeffer, J. 2016. Identifying Platform Effects in Social Media Data. In *International Conference on Web and Social Media*.

Mamykina, L.; Nakikj, D.; and Elhadad, N. 2015. Collective Sensemaking in Online Health Forums. In *Proceedings of CHI*.

Manzer, J. L.; and Bell, A. V. 2022. "Did I Choose a Birth Control Method Yet?": Health Care and Women's Contraceptive Decision-Making. *Qualitative Health Research*.

Merz, A. A.; Gutiérrez-Sacristán, A.; Bartz, D.; Williams, N. E.; Ojo, A.; Schaefer, K. M.; Huang, M.; Li, C. Y.; Sandoval, R. S.; Ye, S.; et al. 2020. Population attitudes toward contraceptive methods over time on a social media platform. *American Journal of Obstetrics and Gynecology*.

Mimno, D.; Wallach, H.; Talley, E.; Leenders, M.; and McCallum, A. 2011. Optimizing Semantic Coherence in Topic Models. In Barzilay, R.; and Johnson, M., eds., *Proceedings of EMNLP*.

Mondal, M.; Liu, Y.; Viswanath, B.; Gummadi, K. P.; and Mislove, A. 2014. Understanding and specifying social access control lists. In *Symposium On Usable Privacy and Security*.

Nahata, L.; Chelvakumar, G.; and Leibowitz, S. 2017. Gender-affirming pharmacological interventions for youth with gender dysphoria: when treatment guidelines are not enough. *Annals of Pharmacotherapy*.

Nelson, A. L.; Cohen, S.; Galitsky, A.; Hathaway, M.; Kappus, D.; Kerolous, M.; Patel, K.; and Dominguez, L. 2017. Women's perceptions and treatment patterns related to contraception: results of a survey of US women. *Contraception*.

Nelson, L. K. 2020. Computational grounded theory: A methodological framework. *Sociological Methods & Research*.

Nobles, A.; Dreisbach, C.; Keim-Malpass, J.; and Barnes, L. 2018. "Is this an STD? Please help!": Online information seeking for sexually transmitted diseases on Reddit. In *Proceedings of ICWSM*.

Nobles, A. L.; Dredze, M.; and Ayers, J. W. 2019. "Repeal and replace": increased demand for intrauterine devices following the 2016 presidential election. *Contraception*.

Nobles, A. L.; Leas, E. C.; Dredze, M.; and Ayers, J. W. 2020. Examining Peer-to-Peer and Patient-Provider Interactions on a Social Media Community Facilitating Ask the Doctor Services. In *Proceedings of ICWSM*.

Olteanu, A.; Castillo, C.; Diaz, F.; and Kıcıman, E. 2018. Social Data: Biases, Methodological Pitfalls, and Ethical Boundaries. *Frontiers in Big Data*, 2.

Patel, D.; Blandford, A.; Warner, M.; Shawe, J.; and Stephenson, J. 2019. "I feel like only half a man" Online Forums as a Resource for Finding a "New Normal" for Men Experiencing Fertility Issues. In *Proceedings of CSCW*.

Pei, J.; and Jurgens, D. 2020. Quantifying Intimacy in Language. In *Proceedings of EMNLP*.

Perrin, A.; and Anderson, M. 2019. Share of US adults using social media, including Facebook, is mostly unchanged since 2018. *Pew Research Center*.

Pfeffer, J.; Mooseder, A.; Hammer, L.; Stritzel, O.; and Garcia, D. 2022. This Sample seems to be good enough! Assessing Coverage and Temporal Reliability of Twitter's Academic API. *ArXiv*, abs/2204.02290.

Pillarisetti, L.; Hyzer, R. H.; Kerr, B. R.; and Moreno, M. A. 2022. 109. Exploring The Discussion Of Nexplanon (Etonogestrel Birth Control Implant) On Twitter. *Journal of Adolescent Health*.

Pirolli, P.; and Card, S. 2005. The sensemaking process and leverage points for analyst technology as identified through cognitive task analysis. In *Proceedings of International Conference on Intelligence Analysis*.

Rivas, R.; Sadah, S. A.; Guo, Y.; and Hristidis, V. 2020. Classification of Health-Related Social Media Posts: Evaluation of Post Content–Classifier Models and Analysis of User Demographics. *JMIR Public Health and Surveillance*.

Röder, M.; Both, A.; and Hinneburg, A. 2015. Exploring the Space of Topic Coherence Measures. In *Proceedings of WSDM*.

Russo, J. A.; Parisi, S. M.; Kukla, K.; and Schwarz, E. B. 2013. Women's information-seeking behavior after receiving contraceptive versus noncontraceptive prescriptions. *Contraception*.

Samulowitz, A.; Gremyr, I.; Eriksson, E. M.; and Hensing, G. 2018. "Brave Men" and "Emotional Women": A Theory-Guided Literature Review on Gender Bias in Health Care and Gendered Norms towards Patients with Chronic Pain. *Pain Research & Management*.

Sannon, S.; Murnane, E. L.; Bazarova, N. N.; and Gay, G. 2019. "I Was Really, Really Nervous Posting It": Communicating about Invisible Chronic Illnesses across Social Media Platforms. In *Proceedings of CHI*.

Scarry, E. 1987. *The body in pain: The making and unmaking of the world*. Oxford University Press, USA.

Schindler, A. E. 2013. Non-contraceptive benefits of oral hormonal contraceptives. *International Journal of Endocrinology and Metabolism*.

Schofield, A.; Magnusson, M.; and Mimno, D. 2017. Pulling Out the Stops: Rethinking Stopword Removal for Topic Models. In *Proceedings of EACL*.

Schofield, A.; and Mimno, D. 2016. Comparing Apples to Apple: The Effects of Stemmers on Topic Models. *TACL*.

Schofield, A.; Thompson, L.; and Mimno, D. 2017. Quantifying the Effects of Text Duplication on Semantic Models. In *Proceedings of EMNLP*.

Tangherlini, T. R. 2000. Heroes and lies: Storytelling tactics among paramedics. *Folklore*.

Weick, K. E.; Sutcliffe, K. M.; and Obstfeld, D. 2005. Organizing and the process of sensemaking. *Organization Science*.

Yang, D.; Kraut, R. E.; Smith, T.; Mayfield, E.; and Jurafsky, D. 2019. Seekers, providers, welcomers, and storytellers: Modeling social roles in online health communities. In *Proceedings of CHI*.

Yee, L.; and Simon, M. 2010. The role of the social network in contraceptive decision-making among young, African American and Latina women. *Journal of Adolescent Health*.

Yoost, J. 2014. Understanding benefits and addressing misperceptions and barriers to intrauterine device access among populations in the United States. *Patient Preference and Adherence*.

Young, A. L.; and Miller, A. D. 2019. "This Girl is on Fire": Sensemaking in an Online Health Community for Vulvodynia. In *Proceedings of CHI*.

Zhang, R.; N. Bazarova, N.; and Reddy, M. 2021. Distress Disclosure across Social Media Platforms during the COVID-19 Pandemic: Untangling the Effects of Platforms, Affordances, and Audiences. In *Proceedings of CHI*.

Paper Checklist

1. For most authors...
   (a) Would answering this research question advance science without violating social contracts, such as violating privacy norms, perpetuating unfair profiling, exacerbating the socio-economic divide, or implying disrespect to societies or cultures? Yes, see §Broader Impacts, Ethics, and Limitations
   (b) Do your main claims in the abstract and introduction accurately reflect the paper's contributions and scope? Yes
   (c) Do you clarify how the proposed methodological approach is appropriate for the claims made? Yes, see §Methods
   (d) Do you clarify what are possible artifacts in the data used, given population-specific distributions? Yes, see §Data Limitations
   (e) Did you describe the limitations of your work? Yes, we address general data limitations as well as the demographic and inherent limitations of social media analysis in general in the §Data Limitations, §Limitations, and §Broader Impacts, Ethics, and Limitations sections
   (f) Did you discuss any potential negative societal impacts of your work? Yes, see §Broader Impacts, Ethics, and Limitations
   (g) Did you discuss any potential misuse of your work? Yes, see §Broader Impacts, Ethics, and Limitations.
   (h) Did you describe steps taken to prevent or mitigate potential negative outcomes of the research, such as data and model documentation, data anonymization, responsible release, access control, and the reproducibility of findings? Yes, see §Broader Impacts, Ethics, and Limitations
   (i) Have you read the ethics review guidelines and ensured that your paper conforms to them? Yes
2. Additionally, if your study involves hypotheses testing...
   (a) Did you clearly state the assumptions underlying all theoretical results? Yes
   (b) Have you provided justifications for all theoretical results? Yes
   (c) Did you discuss competing hypotheses or theories that might challenge or complement your theoretical results? Yes, see §Discussion
   (d) Have you considered alternative mechanisms or explanations that might account for the same outcomes observed in your study? Yes, see §Discussion
   (e) Did you address potential biases or limitations in your theoretical framework? Yes
   (f) Have you related your theoretical results to the existing literature in social science? Yes, see §Related Work
   (g) Did you discuss the implications of your theoretical results for policy, practice, or further research in the social science domain? Yes, see §Discussion and §Broader Impacts, Ethics, and Limitations
3. Additionally, if you are including theoretical proofs...
   (a) Did you state the full set of assumptions of all theoretical results? NA
   (b) Did you include complete proofs of all theoretical results? NA
4. Additionally, if you ran machine learning experiments...
   (a) Did you include the code, data, and instructions needed to reproduce the main experimental results (either in the supplemental material or as a URL)? We provide the lexica for the side effects and contraceptive methods at a Github repository linked in the main paper
   (b) Did you specify all the training details (e.g., data splits, hyperparameters, how they were chosen)? Yes
   (c) Did you report error bars (e.g., with respect to the random seed after running experiments multiple times)? NA
   (d) Did you include the total amount of compute and the type of resources used (e.g., type of GPUs, internal cluster, or cloud provider)? NA
   (e) Do you justify how the proposed evaluation is sufficient and appropriate to the claims made? Yes, see §Methods (Lexicon Evaluation, Model Evaluation)
   (f) Do you discuss what is "the cost" of misclassification and fault (in)tolerance? No
5. Additionally, if you are using existing assets (e.g., code, data, models) or curating/releasing new assets...
   (a) If your work uses existing assets, did you cite the creators? NA
   (b) Did you mention the license of the assets? NA
   (c) Did you include any new assets in the supplemental material or as a URL? Yes
   (d) Did you discuss whether and how consent was obtained from people whose data you're using/curating? Yes
   (e) Did you discuss whether the data you are using/curating contains personally identifiable information or offensive content? Yes
   (f) If you are curating or releasing new datasets, did you discuss how you intend to make your datasets FAIR? NA
   (g) If you are curating or releasing new datasets, did you create a Datasheet for the Dataset? NA
6. Additionally, if you used crowdsourcing or conducted research with human subjects...
   (a) Did you include the full text of instructions given to participants and screenshots? NA
   (b) Did you describe any potential participant risks, with mentions of Institutional Review Board (IRB) approvals? NA
   (c) Did you include the estimated hourly wage paid to participants and the total amount spent on participant compensation? NA
   (d) Did you discuss how data is stored, shared, and deidentified? NA