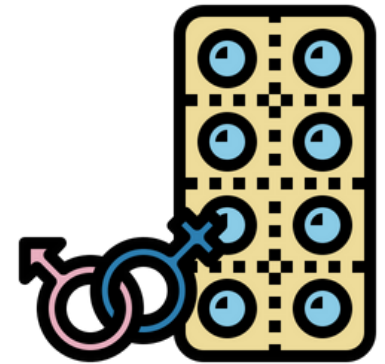# Exploring Language Used in Posts on r/birthcontrol: Case Study Using Data From Reddit Posts and Natural Language Processing to Advance Contraception Research

Elizabeth Pleasants; Julia Holmes Ryan; Cheng Ren; Ndola Prata; Anu Manchikanti Gomez; Cassondra Marshall

Full publication link

Background

The ability to access and choose a contraceptive method is vital to reproductive autonomy. In the United States over the last decade, the majority (about 65%) of reproductive-age people have used some form of birth control [1-3]. Users may choose between daily hormonal contraceptives ("the pill") which has been popular for decades, or long-acting reversible contraception (LARC) methods, like implants and intrauterine devices (IUD), which have become more common in recent years. People's experiences with each method vary, which is why individual preferences are important to the process of making informed contraceptive choices.

R/birthcontrol is an online community on Reddit where people can post about their experiences, ask questions, and seek support from peers related to birth control. **This study analyzed posts on r/birthcontrol to understand the needs and experiences of contraceptive users.**
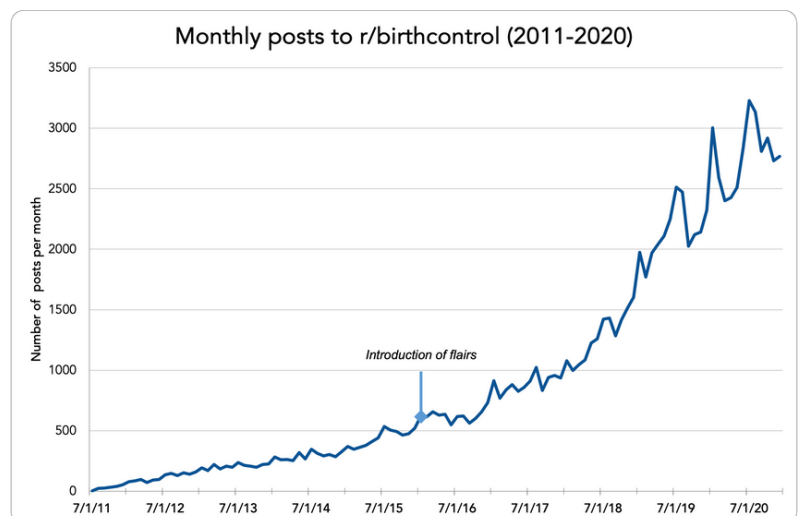
Methods

From the creation of r/birthcontrol on July 21, 2011 to the start date of analysis on December 31, 2020, we looked at the number of posts, their length, and the use of "flairs" (or content tags) to understand people's use of this subreddit.* Part of the analysis focused on popular posts in the subreddit.** A Natural Language Processing tool was used to find the words and phrases people often included in their posts.***

Key Findings

R/birthcontrol is commonly and increasingly used to post about contraception.

Reproductive-aged people are increasingly using online communities. These platforms offer easy access to contraceptive information from peers, specifically about experiences using different contraceptives.
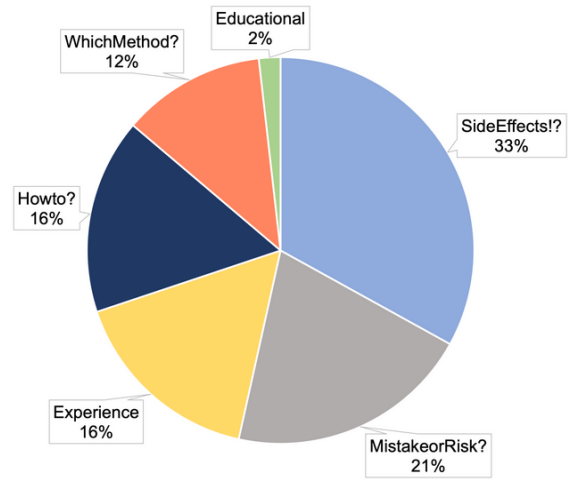
The total number of posts within the study period was 105,485. Since the creation of r/birthcontrol, the number of monthly posts increased over time, especially over the last 5 years. In July 2020, the number of posts within that month was almost 3500.

Monthly posts to r/birthcontrol (2011-2020)

Many members of r/birthcontrol used it to discuss their experiences with contraceptive side effects.

"SideEffects!?" was the most used flair with about 40% of the total posts, or 27,530 posts in the study period. In posts about side effects, people often wrote about IUDs, birth control pills, and specific side effects like sex drive, weight gain, and mood swings. This finding may indicate that people's need for information and support relating to side effects may not be met by current clinical support regarding side effects.

Percent of Posts with Each Flair From 2016-2020 (n = 68,824 posts)

- Educational 2%
- WhichMethod? 12%
- SideEffects!? 33%
- Howto? 16%
- Experience 16%
- MistakeorRisk? 21%

Impacts and Opportunities

- **Clinicians**: Current contraceptive counseling leaves people with gaps in information and support, and online communities like r/birthcontrol are helping to fill those gaps. Real-time information from online communities provides clinicians with insight into the current needs for contraceptive support.

- **Researchers and advocates**: Real-time, easily accessible information about what contraceptive methods people are interested in is very important, particularly in light of increasing restrictions on reproductive health care in the United States. In the future, researchers should study this community more closely, especially related to how misconceptions and misinformation are shared.

- **Contraceptive users**: Peers are an important source of information as people make decisions about contraception, and Reddit is one place where people can connect with peers around this topic. People should critically engage with online sources of contraceptive information while prioritizing their personal needs and experiences.

Endnotes:

*We extracted data from r/birthcontrol for this analysis using the PushShift Reddit Application programming interface (API). PushShift is a data storage and organization platform for social networking sites that started to collect Reddit data in 2015.

**'Popular' posts were defined as having more than 9 comments and a score (upvotes minus downvotes) of 3 or higher.

***We used TF-IDF, or Term Frequency-Inverse Document Frequency. TF (term frequency) is the number of times the term appears in a document. IDF (inverse document frequency) measures how important a word is across the entire collection of documents.

References:

1. Daniels K, Abma JC. Current contraceptive status among women aged 15-49: United States, 2017-2019. NCHS Data Brief 2020 Oct(388):1-8 [Medline]
2. Kavanaugh ML, Jerman J. Contraceptive method use in the United States: trends and characteristics between 2008, 2012 and 2014. Contraception 2018 Jan;97(1):14-21 [https://linkinghub.elsevier.com/retrieve/pii/S0010-7824(17)30478-X] [CrossRef] [Medline]
3. Kavanaugh ML, Pliskin E. Use of contraception among reproductive-aged women in the United States, 2014 and 2016. F S Rep 2020 Sep;1(2):83-93 [https://linkinghub.elsevier.com/retrieve/pii/S2666-3341(20)30038-6] [CrossRef] [Medline]