

Detecting sub-populations in online health communities: a mixed methods exploration of breastfeeding messages in BabyCenter

Calla Glavin Beauregard¹², Parisa Suchdev¹²³, Ashley M. A. Fehr¹², Isabelle T. Smith¹³, Tabia Tanzin Prama¹², Julia Witte Zimmerman¹²³, Carter Ward³, Juniper Lovato¹³⁴, Christopher M. Danforth¹²⁵, Peter Sheridan Dodds¹²⁴⁶

INTRODUCTION & AIM

Parental stress is a nationwide health crisis according to the U.S. Surgeon General's 2024 advisory; expecting parents are also unlikely to share their concerns with providers and report frequent mistreatment [1]. To allay stress, expecting parents seek advice and share experiences in a variety of venues, from in-person birth education classes and parenting groups to virtual communities, for example, BabyCenter, a moderated online forum community with over 4 million members in the United States alone.

In this study, we aim to understand how parents talk about pregnancy, birth, and parenting by analyzing 5.43M posts and comments from the April 2017--January 2024 cohort of 331,843 BabyCenter "birth club" users (that is, users who participate in due date forums or "birth clubs" based on their babies' due dates).

METHOD

We use a combination of word rank and ensemble topic modeling to identify relevant breastfeeding materials and trends across Birth Clubs. We employ BERTopic to locate breastfeeding threads and Latent Dirichlet Allocation (LDA) to summarize themes. Iterating between these two topic modeling methods, we filter posts and users by breastfeeding content contribution. We then identify the most common topics in the full population, as well as within the subset breastfeeding population. We then explore trends among all user content versus those who posted in threads related to breastfeeding topics using LDA. Additionally, we further explore trends by examining word usage and rank over time within the birth clubs.



Figure 1. Normalized Topic Proportions by BabyCenter Birth Club. These panels represent the topic proportions ranked by size of the top 20 topics in the full birth club dataset (left) and of all posts and comments by users who posted in dominant breastfeeding topics (right). Shared topics across both datasets have the same color. While "Family/Work" appears across both groups, "anxiety" only appears as a dominant topic for the breastfeeding subpopulation.

RESULTS & DISCUSSION

Analyzing time series of word rank, we find that posts and comments containing anxiety-related terms increased steadily from April 2017 to January 2024. We find that the topic of sleep dominates in content generated by the breastfeeding population, as well anxiety-related and work/daycare topics that are not predominant in the full BabyCenter birth club database.

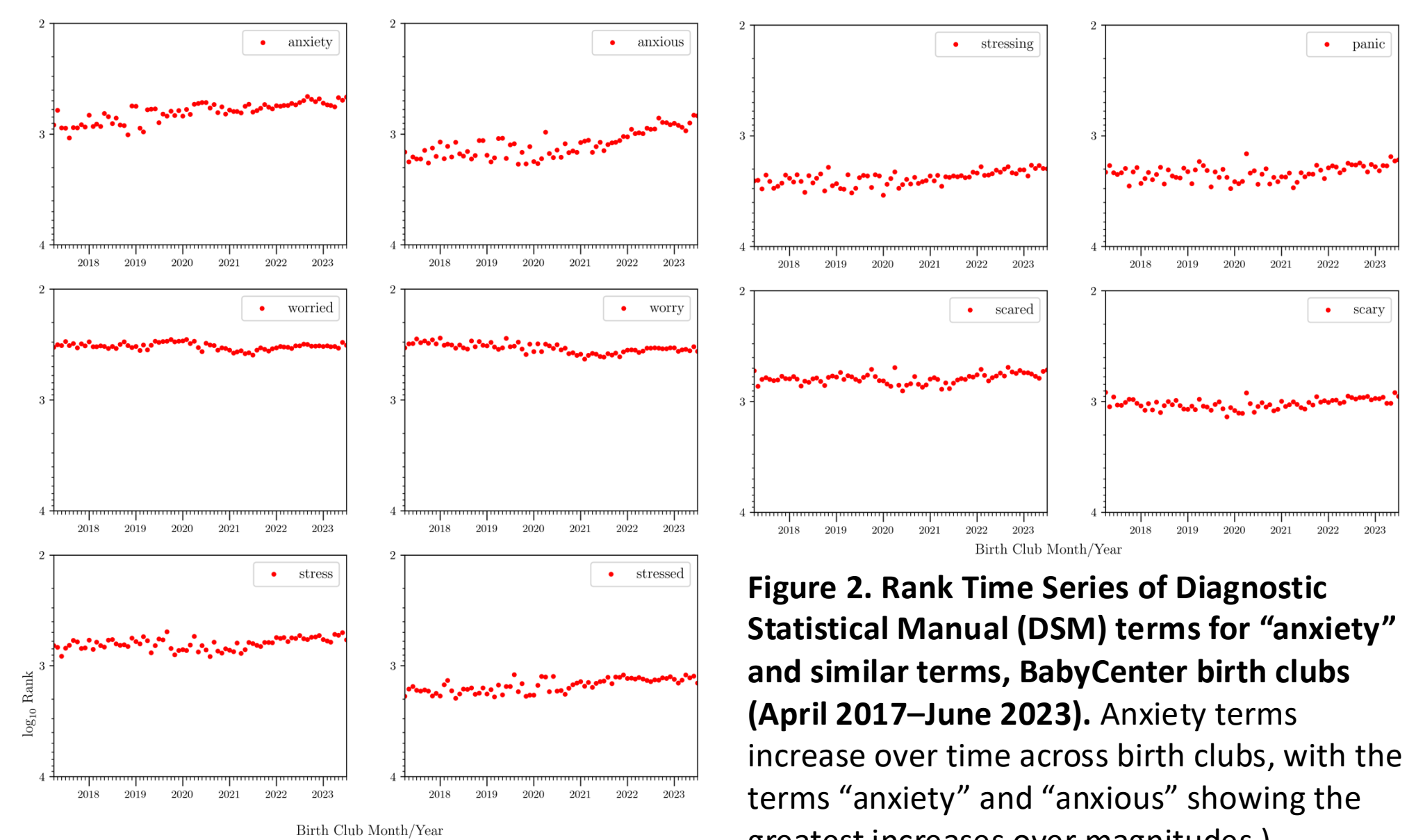


Figure 2. Rank Time Series of Diagnostic Statistical Manual (DSM) terms for "anxiety" and similar terms, BabyCenter birth clubs (April 2017–June 2023). Anxiety terms increase over time across birth clubs, with the terms "anxiety" and "anxious" showing the greatest increases over magnitudes.)

CONCLUSION

Considering the limitations of traditional surveys (measurement, coverage, and response bias) [2], as well as reluctance by birthing people to relate issues to their providers [1], exploring online communities for risk and protective factors for various conditions is an effective time- and cost-saving venture. Our work makes a first step in identifying an important sub-population of breastfeeding people.

FUTURE WORK / REFERENCES

- Centers for Disease Control and Prevention. (2023). *Mistreatment during maternity care*. <https://www.cdc.gov/vitalsigns/respectful-maternity-care/index.html>
- Coughlan, M., Cronin, P., & Ryan, F. (2009). Survey research: Process and limitations. *International Journal of Therapy and Rehabilitation*, 16(1), 9–15.

AUTHOR AFFILIATIONS

¹ Vermont Complex Systems Institute, University of Vermont, Burlington, VT 05405, USA

² Computational Story Lab, University of Vermont, Burlington, VT 05405, USA

³ Computational Ethics Lab, University of Vermont, Burlington, VT 05405, USA

⁴ Department of Computer Science, University of Vermont, Burlington, VT 05405, USA

⁵ Department of Mathematics and Statistics, University of Vermont, Burlington, VT 05405, USA

⁶ Santa Fe Institute, 1399 Hyde Park Rd, Santa Fe, NM 87501, USA