

Using Twitter Data to Study Adverse Pregnancy Outcomes

Ari Z. Klein¹, Davy Weissenbacher¹, Lisa D. Levine², Graciela Gonzalez-Hernandez³

¹ Department of Biostatistics, Epidemiology, and Informatics, University of Pennsylvania

² Department of Obstetrics and Gynecology, University of Pennsylvania

³ Department of Computational Biomedicine, Cedars-Sinai Medical Center

ariklein@penmedicine.upenn.edu, graciela.gonzalez@cshs.org



Background

- In the U.S., 17% of pregnancies end in fetal loss: miscarriage or stillbirth. Preterm birth affects 10% of live births in the U.S. and is the leading cause of neonatal death globally. Preterm birth with low birthweight is the second leading cause of infant mortality in the U.S. The causes of these adverse pregnancy outcomes are largely unknown.

Objectives

- In recent work¹, we used public Twitter data in a case-control study to assess associations between medication exposure during pregnancy and birth defects^{2,3}. The objective of the present study was to develop an automated natural language processing pipeline (NLP) to identify Twitter users reporting adverse pregnancy outcomes in addition to birth defects, enabling the broader use of Twitter data for clinical epidemiology and exploring patient experiences and perceptions.

Conclusions

- Our pipeline enables a unique opportunity to observe early pregnancy—a critical time in fetal development, despite that most people have yet to receive prenatal care. For example, identifying users reporting a miscarriage enables the use of their tweets during early pregnancy to explore risk factors beyond the limited ones for which there are records. By providing opportunities to explore risk factors and gain insights from patients, Twitter data can have clinical value as a complementary resource for informing prenatal care.

Methods

- Data collection.** We developed handwritten regular expressions to search for tweets that mention miscarriage, stillbirth, preterm birth/premature labor, low birthweight, neonatal intensive care unit (NICU) admission, or fetal/infant loss, among more than 400 million tweets that were posted by more than 100 million users who have announced their pregnancy on Twitter⁴.
- Annotation.** Two annotators distinguished tweets that self-report an adverse pregnancy outcome (examples below) from tweets that merely mention an outcome, among 8,109—one random tweet per user—of the 22,912 tweets that matched the regular expressions. Inter-annotator agreement was 0.90 (Cohen's kappa).

It's hard to believe I should be about halfway through my pregnancy right now... #miscarriage

Stress can be vary dangerous! I just gave birth to my 22 week old still born baby girl, so now I've lost it all!

Really? My son was born at 36 weeks because I had preeclampsia! That's crazy!

On December 13th we welcomed our sweet baby girl! Weighs 4lbs 12oz and 18 and a half inches long

I never knew how hard it would be to leave my baby in the NICU over night. #newmomlife

Thinking about my baby today because today was your #duedate #babyloss

- Classification.** We split the 8,109 annotated tweets into 80% and 20% random sets to train and evaluate binary classifiers.

Results

- Annotation.** Among the 8,109 annotated tweets, 3,653 (45%) self-report an adverse pregnancy outcome, including 1,632 that mention miscarriage, 119 stillbirth, 749 preterm birth/premature labor, 217 low birthweight, 558 NICU admission, and 458 fetal/infant loss.
- Classification.** A deep neural network classifier based on bidirectional encoder representations from transformers (BERT)⁴ achieved an overall F_1 -score of 0.88 (precision = 0.87, recall = 0.89) for identifying tweets that self-report an adverse pregnancy outcome, with an F_1 -score of at least 0.82 for each of the specific adverse pregnancy outcomes.

Outcome	F_1 -score	Precision	Recall
Miscarriage	0.88	0.86	0.90
Stillbirth	0.86	0.85	0.88
Preterm birth	0.82	0.84	0.81
Low birthweight	0.88	0.87	0.89
NICU admission	0.92	0.88	0.95
Fetal loss	0.93	0.93	0.93

- Shared task.** Results from the 2021 Social Media Mining for Health Applications (SMM4H) shared tasks⁶ demonstrate that state-of-the-art pretrained transformer models can further improve the overall F_1 -score to 0.93.

Acknowledgments

- This work was supported by the National Library of Medicine (grant R01LM011176). The authors thank Karen O'Connor and Alexis Upshur for annotating tweets.

- Golder S, Chiuve S, Weissenbacher D, Klein A, O'Connor K, Bland M, Malin M, Bhattacharya M, Scarazzini LJ, Gonzalez-Hernandez G. Pharmacoepidemiologic evaluation of birth defects from health-related postings in social media during pregnancy. *Drug Saf.* 2019;42(3):389-400.
- Klein AZ, Sarker A, Cai H, Weissenbacher D, Gonzalez-Hernandez G. Social media mining for birth defects research: a rule-based, bootstrapping approach to collecting data for rare health-related events on Twitter. *J Biomed Inform.* 2018;87:68-78.
- Klein AZ, Sarker A, Weissenbacher D, Gonzalez-Hernandez G. Towards scaling Twitter for digital epidemiology of birth defects. *NPJ Digit Med.* 2019;2:96.
- Sarker A, Chandrashekar P, Magge A, Cai H, Klein A, Gonzalez G. Discovering cohorts of pregnant women from social media for safety surveillance and analysis. *J Med Internet Res.* 2017;19(10):e361.
- Devlin J, Cheng M, Lee K, Toutanova K. BERT: pre-training of deep bidirectional transformers for language understanding. In: Proceedings of the 17th Annual Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (NAACL-HLT); June 2-7, 2019; Minneapolis, MN p. 4171-4186.
- Magge A, Klein AZ, Miranda-Escalada A, Al-Garadi MA, Alimova I, Miftahutdinov Z, Farre E, Lopez SL, Flores I, O'Connor K, Weissenbacher D, Tutubalina E, Sarker A, Banda J, Krallinger M, Gonzalez-Hernandez G. Overview of the sixth Social Media Mining for Health Applications (#SMM4H) shared tasks at NAACL 2021. In: Proceedings of #SMM4H '21: The Sixth Social Media Mining for Health Applications Workshop & Shared Task; June 10, 2021; Virtual p 21-32.

Scan to download the full paper:

