Shatte, A. B. R., Hutchinson, D. M., Fuller-Tyszkiewicz, M., & Teague, S. J. (2020). Social Media Markers to Identify Fathers at Risk of Postpartum Depression: A Machine Learning Approach. Cyberpsychology*, Behavior and Social Networking, 23*(9), 611–618.

https://doi.org/10.1089/cyber.2019.0746

Social media markers to identify fathers at risk of postpartum depression

Adrian B. R. Shatte[1*]

Delyse M. Hutchinson[2345]

Matthew Fuller-Tyszkiewicz[2]

Samantha J. Teague[2]

[1] School of Science, Engineering & Information Technology, Federation University, Melbourne, Australia

[2] Centre for Social and Early Emotional Development, School of Psychology, Deakin University, Geelong, Australia

[3] Murdoch Children's Research Institute, Centre for Adolescent Health, Royal Children's Hospital, Melbourne, Australia

[4] Department of Paediatrics, University of Melbourne, Royal Children's Hospital, Melbourne, Australia

[5] University of New South Wales, National Drug and Alcohol Research Centre, Sydney, Australia

*Corresponding author:

Dr Adrian Shatte

Postal address: Federation University, 100 Clyde Road, Berwick VIC 3806

Phone: +613 51226801

Email: a.shatte@federation.edu.au

ABSTRACT

Postpartum depression is a significant mental health issue in mothers and fathers alike; yet at-risk fathers often come to the attention of healthcare professionals late due to low awareness of symptoms and reluctance to seek help. The present study aimed to examine whether passive social media markers are effective for identifying fathers at risk of postpartum depression. We collected 67,796 Reddit posts from 365 fathers, spanning a six-month period around the birth of their child. A list of 'at risk' words was developed in collaboration with a perinatal mental health expert. Postpartum depression was assessed by evaluating the change in fathers' use of words indicating depressive symptomatology after childbirth. Predictive models were developed as a series of Support Vector Machine (SVM) classifiers using behaviour, emotion, linguistic style and discussion topics as features. The performance of these classifiers indicates that fathers at risk of postpartum depression can be predicted from their prepartum data alone. Overall, the best performing model used discussion topic features only with a recall score of 0.82. These findings could assist in the development of support and intervention tools for fathers during the prepartum period, with specific applicability to personalized and preventative support tools for at-risk fathers.

1. INTRODUCTION

The mental health of fathers across the perinatal period is increasingly recognised as a major public health issue. The estimated rates of depression for fathers in the perinatal period range from 2.1-12% [1], with meta-analyses estimating the overall prevalence rate at approximately 8.4-10.1% [2,3]. Risk factors for depression in fathers include having a partner with depression, low emotional support from one's partner, high negative affect, low dyadic adjustment, and high gender-role stress [2–4]. Perinatal depression in fathers has been linked to poorer mental health in mothers and inter-partner relational and parenting problems [2]. Further, paternal depression is associated with an increased risk of behavioural and emotional difficulties in offspring, similar in magnitude to that associated with maternal psychiatric disorder [5]. Notably, fathers' depression in the perinatal period is associated with higher rates of behaviour and emotional problems in preschool-aged children, psychiatric disorder in school-aged children, and depression in adolescence [5,6].

Despite growing evidence of the role of fathers' mental health in family wellbeing, scant research evidence exists on how best to identify, monitor, and target interventions to fathers presenting with postpartum depression. Reviews to date have consistently noted the paucity of available mental health-focused research on fathers, especially in comparison with research on mothers [2,7]. Evidence suggests that fathers are a difficult group to engage in mental health programs due to the lack of treatment options available, the perceived indifference of fathers in engaging in perinatal care, and fathers' own lack of help-seeking [8,9]. Although fathers may be difficult to engage in traditional healthcare settings, increasing evidence suggests that fathers may be engaging in mental health information seeking digitally via social media platforms. Notably, perinatal healthcare services have reported increased engagement with fathers when offering their services digitally [10,11]; researchers have found increased engagement in fathers through social media when compared to traditional recruitment techniques [12]; and fathers themselves report using the internet as a key source of information on perinatal health issues [13], with rich qualitative studies describing fathers use of social media as a tool for social support in a broad range of fathering contexts [14–17].

Social media may be particularly relevant for research aiming to target depression and other mental health conditions, given the association of high social media use with elevated rates of poor mental health amongst users [18]. Further, social media may also enable the monitoring of fathers at-scale through machine learning approaches that merge sophisticated statistical techniques of large data sets (in this case, text entries on social media) with research in linguistics (e.g., through natural language processing). In brief, natural language processing is the use of computational techniques to make sense of natural language and speech, by converting unstructured text into a structured format for analysis. Natural language processing techniques work particularly well for social media data by identifying the frequency and pairing of different words to formulate an impression of the valence of posts. Such techniques can also be used to extract from social media data the frequency of interactions, linguistic style (e.g. use of pronouns and informal language), emotions, and topics of discussions, to provide a rich profile of fathers. To the best of our knowledge, this novel area of research consists of two studies to date: Teague and Shatte [19] explored the changes in discussion topics as fathers transition across the perinatal period, and Ammari et al. [20] compared parenting roles and identities across discussion forums for mothers and fathers, respectively. Whilst neither of these studies focused on fathers' mental health specifically, both identified that fathers discuss their mental health struggles on the social media site Reddit. This suggests that using natural language processing on father's Reddit content may be a promising method to detect fathers at risk of postnatal depression.

Considerably more research has been conducted using natural language processing techniques on mothers' social media data. Several studies have investigated mothers' postpartum depression using social media data. De Choudry et al. [21] and Morris [22] both identified changes in the posting behaviours of mothers after the birth of their infant on Twitter and Facebook, respectively. Morris[22] identified a sharp decrease in social media use following birth, combined with a change from mostly text-based posts prenatally to multi-media posts postpartum. Further, mothers with postpartum depression were found to use more negative emotional words in posts [22]. De Choudry et al. [21] identified a subset of mothers on Twitter who experienced behavioural changes consistent with depression after announcing the birth of their child, including decreased positive affect and increased negative affect, increased use of first-person pronouns and decreased use of third-person pronouns, and decreases in activation and dominance. This work was then extended to identify prenatal social media indicators of mothers who experience extreme changes in their emotions and social media behaviours after birth, finding that a classification model could identify mothers who would experience extreme changes post-birth with high confidence, yielding an accuracy of 71% [23]. While these studies offer exciting avenues of research for fatherhood researchers, their findings may not directly reflect the experiences of fathers themselves. Ammari et al. [20] identified both similarities and differences in the social media post content of mothers compared to fathers. Further, the presentation of depression itself may differ between men and women, with men typically displaying more substance misuse and risk taking, and poorer impulse control, compared to women [24,25]. Such differences may be reflected in their social media behaviors, including usage patterns and discussion content.

1.1 Objectives

Examining fathers' social media data may enable researchers and health professionals alike to identify and monitor fathers at-risk of poor mental health and offer intervention to a group that are traditionally more difficult to engage in healthcare settings. Such opportunities are particularly pertinent to fathers in the perinatal period, as this is a critical time when the foundations for healthy family functioning and child development are established. Therefore, this study aimed to examine fathers' social media posts in the perinatal period, focusing specifically on changes in behaviour, emotion, linguistic style and discussion topics following the birth of their child. First, a method to identify birth announcements by fathers on social media was developed. Next, data was collected and compared on the prenatal and postnatal posts of fathers including reported behaviour (social media activity and post quality), emotions (negative and positive affect, anger, anxiety and sadness), linguistic style (linguistic dimension, grammar, informal language), discussion topic (categories of forums in which fathers participated on Reddit), and depression symptoms (including mood, cognition, treatment and coping). Finally, a predictive classification model was developed to identify fathers with increased depression symptoms postpartum using their prepartum behaviour, emotions, linguistic style and discussion topics.

2. MATERIALS AND METHODS

2.1 Data

To compare pre and postpartum changes in fathers on social media, we collated a sample of fathers who reported birth events on the forum Reddit. Reddit was chosen due to its popularity (averaging 274 million unique users per month) and its largely male user-base[26,27]. Birth events were searched within a popular discussion group for fathers with over 25,000 subscribers as of January, 2019. Previous research exploring the transition to fatherhood has found that fathers within discussion groups on Reddit announce birth events to the community by describing their "graduation" into fatherhood[19]. Key terms were extracted from the birth announcements category identified by Teague and Shatte[19], described in Table 1, which were then used as search queries for submissions from the fatherhood forum across 2016-2018. The authors of the resultant list of n=610 submissions formed an initial candidate group of fathers. Next, the submissions by the candidate group were independently reviewed by two of the current study's authors to confirm whether the submission was a birth announcement. Agreement between the two reviewers was 98.52%, with disagreements resolved through case conferencing. The final sample consisted of n=365 fathers.

We collated all submissions and comments by the cohort across the full Reddit forum within six months of each father's birth announcement, totaling n=67,796 combined posts and comments. The prepartum phase consisted of submissions and comments in the three months prior to the birth announcement (n=35,833), and the postpartum phase consisted of submissions and comments in the three months after the birth announcement (n=31,963). The six-month pre and postpartum assessment window was chosen to capture fathers' data during mid-pregnancy, when risk of distress is at its highest[1]. Meta-data were extracted from the fathers' posts across several measures, detailed below.

Table 1

List of key terms identified from birth announcements within fatherhood discussion groups on Reddit

graduated; graduation; see you; other side; flipside; see y'all
meet; welcome; delivered; present; world; this is; newest; introducing; born; name; birth; arrive; everyone
oz; lbs; lb; weigh; long; inches
love; beautiful; little; perfect; great
man; boy; girl; tyke
am; pm; /18; yesterday; night; day; today; early; late; weeks
labor; crowning; water broke; c-section; hours; hrs; nicu; induce

2.2 Measures

Measures were developed to characterise fathers' behaviours, emotions, linguistic style, discussion topics, and depression symptomatology in both the pre and postpartum periods.

2.2.1 Behaviours

Fathers' behaviours were considered using four types of measures. First, fathers' engagement with the social media platform was measured through the average number of submissions (i.e., creation of new discussion threads) and comments (i.e., responses to submissions within an existing discussion thread). Second, the quality of fathers' engagement was assessed using the *score* for each submission and comment they contributed. Score is the total number of votes awarded by the reddit community for a submission or comment, calculated by subtracting the community's downvotes from their upvotes. Third, fathers' use of question-centric submissions and comments were evaluated using the number of question marks. Previous research has found that question marks are an adequate heuristic for assessing questioning activities on social media platforms such as Facebook and Twitter [22]. Finally, the total word count contributed by fathers as either submissions or comments.

2.2.2 Emotions

Fathers' emotions were assessed using the percentage of total words from two categories of the psychometrically validated Linguistic Inquiry and Word Count (LIWC) software: positive and negative emotions [28]. The positive emotions category contains 620 words related to positive emotions (e.g., love, nice, sweet), while the negative emotions category contains 744 words related to negative emotions (e.g., hurt, ugly, nasty).

2.2.3 Linguistic Style

Fathers' linguistic style was also assessed using the LIWC software [28]. To extract linguistic features, each fathers' text was evaluated for the percentage of total words from the following linguistic dimensions and other grammar: articles, auxiliary verbs, conjunctions, adverbs, impersonal pronouns, personal pronouns, prepositions, functional words, fillers, assent, negation, certainty, quantifiers. Informal language was also assessed via swear words, "netspeak" (e.g. btw, lol, thx), assent (e.g., agree, OK, yes), nonfluencies (e.g., er, hm, umm), and fillers (e.g., Imean, youknow).

2.2.4 Discussion Topics

Fathers' discussion topics were assessed by categorizing the subreddits that fathers posted submissions or comments to using a community-developed framework. The framework, developed by the reddit community r/TheoryOfReddit, consists of 13 primary categories: gender, gifs, hobbies/occupations, humour, images, lifestyle, porn, technology, videos, and other [29]. These primary categories are further divided into 111 subcategories. This resulted in a range of variables that indicated the number of times each father submitted content within both primary and secondary categories across the pre and postpartum periods.

2.2.5 Depression Symptomatology

Risk of depression was assessed through postpartum changes in fathers' use of depression symptomatology terms using an adapted version of Karmen et al.'s procedure [30]. A lexicon of depression symptomatology terms was developed by reviewing the literature on postpartum depression in fathers. Five systematic reviews on postpartum depression in fathers were identified [3,31–34], containing 328 studies combined. A total of 22 postpartum depression measures were identified across studies, which were then reviewed for key terms of depression symptomatology. Depression symptomatology were grouped according to the ICD-10 symptom groupings, namely: mood, loss of pleasure, appetite, sleep, psychomotor, emotions, cognition, and thoughts of death/suicide. An additional grouping was made labelled "treatment/coping", which included terms relating to treatment and use of drugs and other substances. This initial list of terms, totaling 105 terms, formed our initial lexicon.

The initial lexicon was then expanded by mining synonyms for each term using a thesaurus. A clinical psychologist specializing in perinatal mental health (coauthor REDACTED) then assessed the expanded lexicon to ensure that all included terms captured symptomatology of depression. Any synonyms of depression symptoms that contained dual meanings (e.g., the term "blue" may capture both a mood state symptom of depression and a colour) were excluded from the lexicon. The lexicon was then reduced through a process known as lemmatization; each depression term was reduced to its dictionary form so that the inflected forms of similar words were grouped together, before duplicate terms were removed. For instance, the word 'depression' could have variants such as 'depression' and 'depressed', but were truncated to 'depress' as the common base form, rather than being counted as separate entities. This process resulted in a final depression symptomatology lexicon of 509 terms. A custom LIWC dictionary was created using the lexicon including a global depression category consisting of all terms, and separate categories for each theme of depression symptomatology outlined above. The final dataset therefore contained the percentage of total words used for each depression category.

2.3 Analysis

Analyses were conducted using the *scikit-learn, pandas*, and *scipy* packages for Python. Missing data were accounted for using the method of Multivariate Imputation by Chained Equations (*mice*), which models each feature with missing values as a function of other features using a round-robin approach. Initial exploratory analyses were conducted to assess the overall cohort's change across the perinatal period, and to identify differences between fathers deemed at high or low risk of postpartum depression. First, dependent samples t-tests were conducted to compare the cohort on all measures between the pre and postpartum periods. Next, fathers were categorised into either high or low risk for postpartum depression by the observation of either an increase (high risk) or decrease (low risk) in their mean percentage of global depression symptoms from pre to postpartum. Group differences in global depression symptomatology over time were confirmed using a mixed between-within subjects ANOVA (Wilks λ= 0.97, $F(1, 350) = 11.73$, p <.001, partial η^2=.03). Comparisons between the high and low-risk groups during the pre and postpartum periods were conducted using independent samples t-tests. For all analyses, corrections for multiple testing were not conducted due to the exploratory nature of the research question[35], with actual p values and effect sizes reported.

After distinguishing between the high and low risk groups, the prepartum data was then prepared for inclusion in a predictive model. For each set of behaviour, emotion, linguistic style, and discussion topic measures, daily measurements were obtained per user over the three-month period prior to the detected birth event. This resulted in a time series of data for each user, which were used to compute the following features: (1) *total*: the aggregated totals of each variable for fathers over the prepartum period;

(2) *mean*: the average measure of a variable over the duration of the prepartum period; and, (3) *slope*: the behavioural change of fathers over the prepartum period for each variable, calculated using linear regression. This resulted in 150 total features. In order to reduce the dimensionality of the features for classification, feature selection was performed using Linear Support Vector Classification with the *SelectFromModel* function of *scikit-learn* for Python. The number of features included in the final model was thereby reduced to a subset of 59 features, listed in Supplementary File 1.

After experimenting with several classifiers, it was found that a Support Vector Machine (SVM) with linear kernel produced the best results for predicting an increase in depressive language during the postpartum period. For all analyses, 10-fold cross validation was computed, and the results averaged over 100 randomized experimental iterations. To evaluate the classification performance of the models, precision, recall, F-measure and accuracy of the estimations were calculated. *Precision* indicates the proportion of true positives (i.e., high-risk fathers) out of all positive classifications; *recall* indicates the proportion of true positives out of all true positives and false-negatives (i.e., all high-risk fathers in the cohort); the *F-measure* indicates the balance between the precision and the recall of a model; and *accuracy* is the rate of correct classifications across all fathers in the sample.

2.4 Ethics and Privacy in Social Media Research

Due to the public nature of the data collected in this study, the Federation University Human Research Ethics committee waived ethical review. However, the sensitive nature of the information being collected, and the motivation of users to post in a pseudonymous forum mandates that ethical standards must still be applied. In conducting this work, the research team undertook several steps to minimize the impact on fathers' privacy. First, only publicly available data was included in the study – specifically, public expression by fathers on Reddit. Human exposure to the direct social media data was limited as much as possible, with all analyses conducted by machine on aggregated data. Only birth posts were read by a subset of the author team to confirm that the data signaled the birth of a child. All analyses were conducted at the aggregate-level, removing the ability to re-identify individuals from data such as quotes, usernames, or other information that could be used with additional datasets.

3. RESULTS

3.1 Cohort behaviours across the pre and postpartum periods

The final data consisted of 365 fathers who contributed 3,889 submissions and 63,907 comments across a rolling six-month data window (dependent on birth event post). The mean number of submissions was 5.73 (SD=8.05) in the prepartum period and 5.00 (SD=7.13) in the postpartum period, while the mean number of comments were 92.46 (SD=220.03) and 82.62 (SD=154.22) in the pre and postpartum periods, respectively. The average participant directed most of their submissions and comments to subreddits in the *Other* category in the prepartum period (M=18.03, SD=86.91) and the *Lifestyle* category postpartum (M=15.05, SD=26.78). Mean scores on the depression symptom lexicon were 0.84 (SD=1.03) in the prepartum period and 0.82 (SD=0.52) in the postpartum period, with the emotions and mood depression symptom categories having higher scores in both periods.

Meaningful differences between the pre and postpartum periods are displayed in Table 2. Fathers contributed fewer submissions in the postpartum period and posted fewer questions to the reddit caucus. The number of submissions and comments to Hobby/Occupation subreddits dedicated to guns/combat and photography/film reduced, while submissions and comments to Other subreddits on nostalgia/time and unexpected (subreddits that contain contributions where something odd or unexpected happens) increased. A range of linguistic changes occurred, including an increase in positive emotions and use of third-person singular pronouns (e.g., she, her, him). Use of depression symptom terms did not differ across the pre and postpartum periods, with the exception of terms related to psychomotor symptomatology, which decreased after childbirth.

3.2 Comparisons between high and low risk fathers

Fathers were classified as high and low risk using the direction of change in their global depression symptom terms after childbirth, with high-risk fathers increasing and low-risk fathers decreasing usage in the postpartum period. Meaningful differences between the high and low-risk groups cross-sectionally at the pre and postpartum periods are reported in Table 3. Across both the pre and postpartum periods, fathers in the low-risk group engaged in more discussions on Humour and Lifestyle topics and used a higher proportion of swear and anger words than high-risk fathers. Conversely, fathers in the high-risk group contributed fewer submissions and comments, wrote posts with fewer words, and posed fewer questions to the Reddit caucus, than fathers at low-risk. These behaviours occurred whilst simultaneously increasing their proportion of depression-symptom terms after childbirth. Notably, no differences were observed between groups for depression terms in the postpartum period, although the low-risk group were higher in the prepartum period.

Table 2

Means (M), Standard Deviations (SD) and Comparisons of Fathers' Behaviours, Emotions, Linguistic Style, Discussion Topics and Depression Language in the Prenatal and Postnatal Periods

Variable	Prenatal Period M(SD)	Postnatal Period M(SD)	Comparison $t_{(364)}$	Comparison p	Mean Difference 95% CI lower	Mean Difference 95% CI upper	Cohen's d
Behaviours							
No. Submissions	5.73 (8.05)	5 (7.13)	2.57	.01	0.17	1.29	0.13
No. Questions	2.33 (4.02)	1.93 (3.41)	2.79	<.01*	0.12	0.68	0.15
Emotions							
Positive	3.95 (2.25)	4.29 (2.18)	-2.49	.01	-0.59	-0.07	0.13
Linguistic Style							
Third-person Singular Pronoun	1.1 (1.26)	1.58 (1.59)	-4.91	<.01**	-0.67	-0.29	0.26
Informal Language	1.77 (1.33)	1.91 (1.42)	-1.97	.05	-0.29	0.00	0.10
Nonfluencies	0.22 (0.36)	0.28 (0.47)	-2.19	.03	-0.12	-0.01	0.11
Discussion Topics							
Hobbies/Occupations – Guns and Combat	1.54 (10.74)	0.76 (5.21)	2.04	.04	0.03	1.53	0.11
Hobbies/Occupations – Photography/Film	0.15 (1.11)	0.06 (0.76)	2.32	.02	0.01	0.16	0.13
Other – Nostalgia/Time	0.17 (0.78)	0.33 (1.56)	-2.57	.01	-0.28	-0.04	0.14
Other - Unexpected	0.06 (0.42)	0.14 (0.68)	-2.11	.04	-0.15	-0.01	0.11
Depression Language							
Psychomotor	0.05 (0.12)	0.03 (0.09)	2.85	<.01*	0.01	0.04	0.13

NOTES: * p<.01; ** p<.001

Table 3

Means (M), Standard Deviations (SD) and Comparisons of Fathers' Behaviours, Emotions, Linguistic Style, Discussion Topics and Depression Language in the Low and High Risk Groups

	Low Risk (n = 178)	High Risk (n = 187)	Comparison		Mean Difference 95% CI		Cohen's
Prepartum Variables	**M(SD)**	**M(SD)**	$t_{(df)}$	**p**	**lower**	**upper**	**d**
Behaviours							
No. Submissions	6.91 (9.39)	4.61 (6.34)	2.72 (308.65)	<.01*	.64	3.95	.28
No. Questions	2.81 (4.64)	1.86 (3.28)	2.26 (317.26)	.02	.12	1.78	.24
No. Comments	121.48 (285.41)	64.84 (125.14)	2.43 (240.06)	.02	10.80	102.47	.26
Comment Score	1463.47 (4736.1)	484.12 (1178.7)	2.68 (197.83)	. <.01*	258.97	1699.74	.29
Word Count	4779.53 (11783.32)	2422.42 (4687.59)	2.49 (229.39)	.01	490.42	4223.81	.26
Emotions							
Negative	2.15 (1.51)	1.57 (1.12)	4.19 (363)	<.01**	.31	.86	.44
Anxiety	0.29 (0.27)	0.19 (0.25)	3.38 (363)	<.01**	.04	.14	.38
Anger	0.69 (0.55)	0.55 (0.71)	2.16 (363)	.03	.01	.28	.22
Linguistic Style							
Third-person Singular Pronouns	1.23 (1.42)	0.97 (1.08)	2.04 (363)	.04	.01	.53	.21
Swear Words	0.43 (0.48)	0.31 (0.61)	2.1 (363)	.04	.01	.24	.22
Discussion Topics							
Gifs	1.18 (4.21)	0.44 (2.28)	2.08 (269.48)	.04	.04	1.44	.22
Humour	2.96 (8.54)	1.15 (4.28)	2.54 (257.69)	.01	.41	3.21	.27
Images	3.48 (12.46)	1.05 (2.93)	2.54 (195.65)	.01	.54	4.32	.27
Lifestyle	22.32 (51.16)	11.8 (22.55)	2.52 (240.7)	.01	2.29	18.74	.27
Porn	0.42 (2.21)	0.09 (0.42)	1.99 (189.06)	.05	.00	.67	.21
Videos	0.80 (3.22)	0.24 (1.31)	2.17 (231.61)	.03	.05	1.07	.23
Depression Language							
Global Depression	1.17 (1.35)	0.51 (0.36)	6.35 (200.95)	<.01**	.46	.87	.67
Mood	0.41 (0.51)	0.21 (0.19)	4.94 (223.61)	<.01**	.12	.28	.52

	Low Risk (n = 178)	High Risk (n = 187)	Comparison		Mean Difference 95% CI		Cohen's
Prepartum Variables	**M(SD)**	**M(SD)**	$t_{(df)}$	**p**	**lower**	**upper**	**d**
Loss of Pleasure	0.02 (0.07)	0.01 (0.03)	2.63 (234.41)	<.01*	.00	.03	.19
Psychomotor	0.08 (0.16)	0.03 (0.07)	3.7 (241.74)	<.01**	.02	.07	.41
Emotions	0.75 (1.34)	0.30 (0.26)	4.35 (189.41)	<.01**	.24	.65	.47
Coping	0.07 (0.13)	0.02 (0.06)	4.72 (240.47)	<.01**	.03	.07	.50

	Low Risk (n = 178)	High Risk (n = 187)	Comparison		Mean Difference 95% CI		Cohen's
Postpartum Variables	**M(SD)**	**M(SD)**	$t_{(df)}$	**p**	**lower**	**upper**	**d**
Behaviours							
No. Submissions	5.96 (8.08)	4.10 (5.98)	2.49 (325.53)	.01	.39	3.33	.26
No. Comments	100.23 (161.52)	65.87 (145.4)	2.13 (354.6)	.03	2.68	66.05	.22
Comment Score	1585.44 (6201.84)	614.00 (1612.51)	2.03 (199.73)	.04	25.77	1917.11	.22
Word Count	3689.01 (7112.76)	2410.6 (4927.34)	1.99 (313.41)	.05	12.34	2544.47	.21
Emotions							
Anger	0.67 (0.58)	0.54 (0.60)	2 (363)	.05	.00	.24	.22
Linguistic Style							
Pronoun	15.71 (2.88)	14.92 (2.97)	2.57 (363)	.01	.18	1.39	.27
Quantifiers	2.30 (0.86)	2.49 (1.00)	-2.02 (363)	.04	-.39	.00	.20
Informal Language	2.08 (1.54)	1.75 (1.28)	2.25 (363)	.02	.04	.62	.23
Swear Words	0.44 (0.51)	0.31 (0.48)	2.49 (363)	.01	.03	.23	.26
Netspeak	1.02 (1.26)	0.73 (0.71)	2.72 (363)	<.01*	.08	.50	.28
Discussion Topics							
Humour	2.53 (5.88)	1.37 (3.78)	2.22 (299.39)	.03	.13	2.18	.24
Images	2.86 (8.13)	1.01 (3.07)	2.85 (224.22)	<.01*	.57	3.13	.30
Lifestyle	18.11 (33.78)	12.14 (17.32)	2.11 (261.12)	.04	.39	11.54	.22
Other	19.90 (45.07)	8.63 (20.29)	3.06 (243.25)	<.01*	4.01	18.54	.32
Porn	0.67 (3.80)	0.05 (0.24)	2.18 (178.33)	.03	.06	1.18	.23

NOTES: * p<.01; ** p<.001

3.3 Identifying at-risk fathers from prenatal data

Table 4 reports the performance of the SVM classifiers with different feature categories, with receiver-operator characteristic (ROC) curves displayed in Figure 1. These curves depict the relationship between the true-positives (correctly identified high-risk fathers) and false-positives (low-risk fathers incorrectly classified as high-risk) for each model. The model incorporating all features identified high-risk fathers with 0.67 precision, 0.68 recall, 0.67 F-measure, and an accuracy of 66%. The model that included emotion features alone also showed reasonable performance at classifying high-risk fathers, with an accuracy of 62%. Interestingly, behaviour and discussion topic features alone yielded greater recall, with 0.77 and 0.82 respectively, which may be useful for screening purposes.

Table 4

Mean performance metrics using prepartum features to predict high or low risk of fathers' postpartum depression

Features	Precision	Recall	F-measure	Accuracy
Behaviour	0.58	0.77	0.66	60%
Discussion Topic	0.57	0.82	0.67	59%
Linguistic Style	0.56	0.55	0.55	55%
Emotions	0.64	0.58	0.60	62%
All features	***0.67***	***0.68***	***0.67***	***66%***

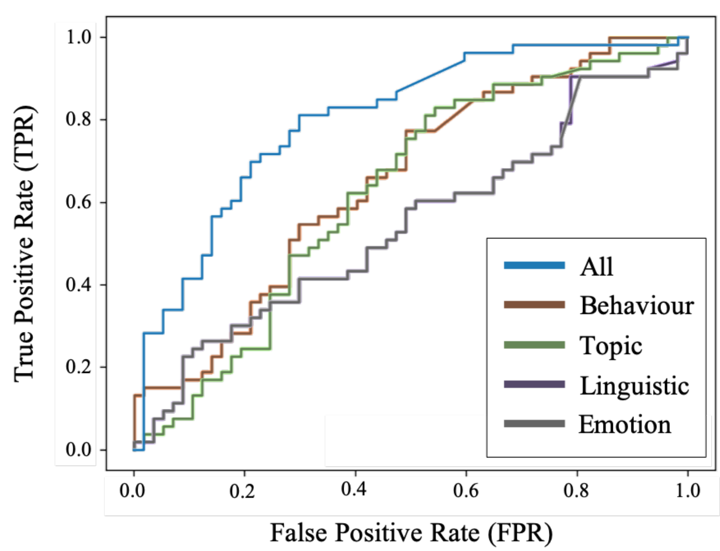

Figure 1: ***ROC curves for prediction using prepartum features. Aggregated trends over measure categories shown.***

4. DISCUSSION

This study aimed to investigate methods for detecting increased depression symptoms in fathers during the postpartum period, using passive social media data gathered during the prepartum period. This was achieved by detecting birth announcements posted by fathers on a popular fatherhood social support subreddit, and then extracting all posts made by the cohort of fathers within a six-month window around the birth event. Metadata were extracted from the posts in terms of behaviour, linguistic style, emotion, and discussion topics, with these data being used to train a predictive machine learning model to detect increased depression symptoms after childbirth. In doing so, this study has demonstrated that social media data capture is a promising passive monitoring method for the detection of men's mental health status across the fatherhood transition. Such research has important implications for the design of online interventions for fathers and for research methodologies used to reach fathers across the perinatal period and beyond.

While the initial exploratory analyses demonstrated that fathers generally appear to be coping well across the perinatal period, a number of social media markers were identified that delineated fathers at high and low-risk for postpartum depression. Comparisons between those who increased their use of depression symptom terms postpartum (high-risk) against those who decreased (low-risk) demonstrated that there are evidently differences in behaviours, emotions, linguistic style and discussion topics across the perinatal period. Notably, high-risk fathers made fewer contributions (submissions, comments, word count) which were rated as lower quality by peers (comment score), and engaged in fewer discussions categorized as Humour, Lifestyle, and Image-related. These observations are consistent with previous research on mothers [21–23,36], and could be explained by symptoms of depression that may discourage social engagement, such as exhaustion and anhedonia. Surprisingly, fathers in the low-risk group were found to have higher anger and use of swear words across both the pre and postpartum periods than the high-risk group. A possible explanation could be that high-risk fathers are less emotionally expressive than low-risk fathers, resulting in greater disclosure from the low-risk group. Depression can be characterized by flat emotional expression to stimuli, which may be reflected on social media by low positive and negative affect, fewer contributions, and lower-quality ratings by peers. Such behaviours were observed in the high-risk group in the current study, making this explanation appealing. A second possible explanation could be that the depression lexicon failed to adequately capture anger as a symptom of depression. Debate in the research literature on the suitability of postpartum depression measures that have typically been adapted from mothers for fathers have raised concerns that such measures do not adequately assess irritability, anger, substance use, and other externalizing depression symptoms [24,25,37]. These measurement limitations could have resulted in our depression lexicon failing to capture the breadth of depression expression in fathers on social media, pointing to the need for improved depression assessment methods for fathers generally.

The results demonstrated that an increase in postpartum depression symptoms displayed in social media text can be detected using fathers' prepartum data alone. Models developed using these features to predict an increase of depression symptom terms in fathers during the postpartum period demonstrated good performance (accuracy = 66%). These results are similar to previous studies of postnatal depression in mothers using other social media platforms (e.g., Twitter [21,23] and Facebook [22,36]), suggesting that both mothers and fathers experiencing postpartum depression exhibit similar digital indicators that can be leveraged for early detection. Interestingly, the behaviour and discussion topic features individually yielded higher recall than the combined model, with 0.77 and 0.82 recall, respectively. Higher recall may be preferred in clinical screening tools to maximise the number of persons in-need of support being identified, despite potentially falsely identifying individuals who do not need such support (i.e., at the expense of higher precision and accuracy) [38]. This suggests that fathers' broad discussion topics may be sufficient for any social media-based intervention and support tools addressing the mental health challenges faced by fathers during the prepartum period. Such interventions could use non-invasive methods to track the characteristics and interactions of fathers over the prenatal period on social media, offering fathers the ability to self-assess and access context-specific support at critical periods [39], e.g., after detecting childbirth.

Importantly, the results demonstrate the potential of social media for longitudinal observation of fathers for their health and wellbeing. Previous studies have demonstrated that engaging fathers in research and clinical settings can be difficult [9,40], indicating that observation through social media data may be particularly useful for this population and other hard-to-reach groups. Social media data have several strengths for longitudinal research, such as ability to capture prospective, in-the-moment data from large groups of people [41]. Anonymous sources such as Reddit may be particularly useful, given that fathers report disclosing more sensitive information on anonymous platforms than identified social media sites (e.g., Facebook) [14]. The current study demonstrates several useful techniques for monitoring fathers' health and wellbeing over time, including measures capable of detecting individual variations in behaviour, emotions, linguistic style, discussion topics, and depression symptoms over time, and methods for identifying significant life events (e.g., childbirth). Further, the collaborative involvement of domain experts in both perinatal mental health and computational social science demonstrates methods to investigate fatherhood using novel, unstructured data whilst navigating construct validity, data biases, and sensitivity [42]. These methods could have applications for other research questions in fathers' health and wellbeing, such as monitoring work-life balance and other mental health conditions such as anxiety and stress.

While there are many positive applications of automatic classification of user groups based on social media data, such as providing hard-to-reach groups with non-invasive monitoring and access to support, there is also the potential for such techniques to be used in a negative manner. For instance, automated techniques could detect significant life events and discussion topics that users may not wish to be identified and collated despite being posted publicly. Such information could be used with malicious intent to garner information about a user for unintended use, potentially exposing the user to unwanted risk. This risk may be mitigated on a forum platform such as Reddit, due to its pseudonymous nature; previous research has indicated that parents may post more sensitive data on Reddit due to the pseudonymous nature of the platform, even creating "throwaway" accounts to reduce the likelihood of certain information being linked to their personal identity [43]. However, similar techniques to those proposed in this study could also be used on other platforms with public data streams that are more likely to be tied to a personal account (e.g., Twitter). As such, researchers and practitioners should be careful when using social media data to ensure that privacy of users is maintained and that the techniques are not exposing users to additional risk. Future research should investigate parents' perceptions of social media data being used by researchers to gather insights on their health and wellbeing. Such research would help to guide the development of ethical research practices for this unique population that balance the interests of both groups.

This study showed promising results for the use of passive social media data to detect an increase in depression symptoms in fathers during the postpartum period; however, it is important to note the research has several limitations. First, results should be interpreted within the context of the Reddit platform that was used to collect data. While Reddit is an internationally accessible website used by a global audience, its primary userbase consists of English-speaking users from North America [26,27]. This may limit the current study's generalizability across cultural contexts; although, previous research has demonstrated that depression symptomatology can be predicted from social media behaviour in cross-cultural groups with general populations (e.g., [44–46]). Compared to other platforms, a key characteristic of Reddit is the pseudonymous nature of user accounts in which no personally identifying information needs to be attached to an account, with users engaging in public discussions with other pseudonymous users who are unlikely to be in their offline social network. It has been argued that the pseudonymous nature of Reddit may actually promote greater disclosure of sensitive parenting topics [14,19,20]; however, future research should further explore the specific nature of Reddit as a platform for fathers mental health disclosure and support.

Additionally, the current study is limited by a lack of ground truth data to verify the research findings. Supplementing social media data with measures completed directly by participants would assist in establishing both users' fatherhood and depression status in the pre and postpartum periods. Fathers are recognized as a group that are difficult to engage in both research and health service contexts [9,40], posing a challenging barrier to the collection of ground truth data. The current study instead used indirect methods to assess fatherhood and depression status. First, fatherhood status was established by membership of a fatherhood forum, consistent with methods used in previous research [47]. Second, depression symptoms were assessed using a novel method proposed by Karmen et al. [30] with involvement by experts in perinatal mental health, which is demonstrated to be the most effective technique in lieu of ground truth data [48]. Nevertheless, future work should therefore aim to address this limitation by combining both fathers online and offline data, including the use of gold-standard mental health measures.

5. CONCLUSION

This study demonstrates the potential for fathers' social media data captured in the prepartum period to predict an increase in depression symptoms during the postpartum period. A range of predictive features were identified, including behaviours, emotions, linguistic style, and discussion topics. The methods presented in this study could assist in the development of support and intervention tools for fathers during the prepartum period, with specific applicability to personalized and preventative support tools for fathers and other hard-to-reach groups.

6. ACKNOWLEDGEMENTS AND DECLARATIONS

6.1 Financial Support

This research received no specific grant from any funding agency, commercial or not-for-profit sectors.

6.2 Conflict of Interest

None.

6.3 Contributions of Authors

AS conceived of the study, participated in its design and coordination, performed the data collection and analysis, interpreted the data, and drafted the manuscript; DH assisted with data collection and analysis, interpretation of results, and helped to draft and revise the manuscript; MFT assisted with interpretation of results, and helped to draft and revise the manuscript; ST conceived of the study, participated in its design and coordination, contributed to the data collection, analysis, and interpretation, and helped to draft and revise the manuscript. All authors read and approved the final manuscript.

REFERENCES

1. Wong O, Nguyen T, Thomas N, Thomson-Salo F, Handrinos D, Judd F. Perinatal mental health: fathers – the (mostly) forgotten parent. *Asia-Pacific Psychiatry*. 2016;8(4):247-255. doi:10.1111/appy.12204

2. Paulson JF, Bazemore SD. Prenatal and postpartum depression in fathers and its association with maternal depression: a meta-analysis. *JAMA*. 2010;303(19):1961-1969.

3. Cameron EE, Sedov ID, Tomfohr-madsen LM. Prevalence of paternal depression in pregnancy and the postpartum : An updated meta-analysis. *J Affect Disord*. 2016;206:189-203. doi:10.1016/j.jad.2016.07.044

4. Morse CA, Buist A, Durkin S. First-time parenthood: influences on pre- and postnatal adjustment in fathers and mothers. *J Psychosom Obstet Gynaecol*. 2000.

5. Ramchandani P, Psychogiou L. Paternal psychiatric disorders and children's psychosocial development. *Lancet*. 2009;374(9690):646-653. doi:10.1016/S0140-6736(09)60238-5

6. Gutierrez-Galve L, Stein A, Hanington L, et al. Association of Maternal and Paternal Depression in the Postnatal Period with Offspring Depression at Age 18 Years. *JAMA Psychiatry*. 2019;76(3):290-296. doi:10.1001/jamapsychiatry.2018.3667

7. Parent J, Forehand R, Pomerantz H, Peisch V, Seehuus M. Father Participation in Child Psychopathology Research. *J Abnorm Child Psychol*. 2017;45(7):1259-1270. doi:10.1007/s10802-016-0254-5

8. Sicouri G, Tully L, Collins D, et al. Toward Father-friendly Parenting Interventions: A Qualitative Study. *Aust New Zeal J Fam Ther*. 2018;39(2):218-231. doi:10.1002/anzf.1307

9. Bateson K, Darwin Z, Galdas P, Rosan C. Engaging fathers: acknowledging the barriers. *J Heal Visit*. 2017;5(3):126-132. doi:10.12968/johv.2017.5.3.126

10. O'Brien AP, McNeil KA, Fletcher R, et al. New fathers' perinatal depression and anxiety—treatment options: an integrative review. *Am J Mens Health*. 2017;11(4):863-876. doi:10.1177/1557988316669047

11. Mackert M, Guadagno M, Lazard A, et al. Engaging men in prenatal health via eHealth: findings from a national survey. *JMIR Pediatr Parent*. 2018;1(2):e7. doi:10.2196/pediatrics.9513

12. Bennetts SK, Hokke S, Crawford S, et al. Using Paid and Free Facebook Methods to Recruit Australian Parents to an Online Survey: An Evaluation. *J Med Internet Res*. 2019;21(3):e11206.

doi:10.2196/11206

13. Johansson M. Sources of information, birth preference, and experiences of childbirth and postnatal care. 2012.

14. Ammari T, Schoenebeck S. Understanding and supporting fathers and fatherhood on social media sites. In: *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems - CHI '15*. ; 2015:1905-1914. doi:10.1145/2702123.2702205

15. Fletcher R, StGeorge J. Heading into fatherhood-nervously: support for fathering from online dads. *Qual Health Res*. 2011;21(8):1101-1114. doi:10.1177/1049732311404903

16. Eriksson H, Salzmann-Erikson M. Supporting a caring fatherhood in cyberspace - an analysis of communication about caring within an online forum for fathers. *Scand J Caring Sci*. 2013;27(1):63-69. doi:10.1111/j.1471-6712.2012.01001.x

17. Kim HN, Wyatt TH, Li X, Gaylord M. Use of social media by fathers of premature infants. *J Perinat Neonatal Nurs*. 2016;30(4):359-366. doi:10.1097/JPN.0000000000000214

18. Yoon S, Kleinman M, Mertz J, Brannick M. Is social network site usage related to depression? A meta-analysis of Facebook–depression relations. *J Affect Disord*. 2019;248(January):65-72. doi:10.1016/j.jad.2019.01.026

19. Teague SJ, Shatte ABR. Exploring the transition to fatherhood: feasibility study using social media and machine learning. *JMIR Pediatr Parent*. 2018;1(2):e12371. doi:10.2196/12371

20. Ammari T, Schoenebeck S, Romero DM. Pseudonymous Parents : Comparing Parenting Roles and Identities on the Mommit and Daddit Subreddits. In: *ACM Conference on Human Factors in Computing Systems (CHI)*. ; 2018:489-502. doi:10.1145/3173574.3174063

21. De Choudhury M, Counts S, Horvitz E. Major life changes and behavioral markers in social media. *CSCW '13*. 2013:1431. doi:10.1145/2441776.2441937

22. Morris MR. Social networking site use by mothers of young children. *CSCW '14*. 2014:1272-1282. doi:10.1145/2531602.2531603

23. De Choudhury M, Counts S, Horvitz E. Predicting postpartum changes in emotion and behavior via social media. 2013:3267. doi:10.1145/2470654.2466447

24. Martin LA, Neighbors HW, Griffith DM. The experience of symptoms of depression in men vs women: Analysis of the national comorbidity survey replication. *JAMA Psychiatry*. 2013;70(10):1100-1106. doi:10.1001/jamapsychiatry.2013.1985

25. Cavanagh A, Wilson CJ, Kavanagh DJ, Caputi P. Differences in the Expression of Symptoms in Men Versus Women with Depression. *Harv Rev Psychiatry*. 2017. doi:10.1097/hrp.0000000000000128

26. Duggan M, Smith A. *6% of Online Adults Are Reddit Users*.; 2013. http://www.webcitation.org/72oqjJxsw.

27. Reddit. Audience and demographics. http://www.webcitation.org/72oqsDpkv. Published 2017.

28. Pennebaker JW, Boyd RL, Jordan K, Blackburn K. *The Development and Psychometric Properties of LIWC2015*. Austin, TX: University of Texas at Austin; 2015. doi:10.1068/d010163

29. r/TheoryofReddit. Subreddit Map.

30. Karmen C, Hsiung RC, Wetter T. Screening internet forum participants for depression symptoms by assembling and enhancing multiple NLP methods. *Comput Methods Programs Biomed*. 2015;120(1):27-36. doi:10.1016/j.cmpb.2015.03.008

31. Paulson JF, Bazemore SD. Prenatal and postpartum depression in fathers and its association with maternal depression: a meta-analysis. *JAMA*. 2010;303(19):1961-1969. doi:10.1001/jama.2010.605

32. Edward K, Castle D, Mills C, Davis L, Casey J, Education DN. An Integrative Review of Paternal Depression. 2015. doi:10.1177/1557988314526614

33. Wilson S, Durbin CE. Effects of paternal depression on fathers ' parenting behaviors: A meta-analytic review. *Clin Psychol Rev*. 2010;30(2):167-180. doi:10.1016/j.cpr.2009.10.007

34. Pilkington PD, Milne LC, Cairns KE, Lewis J, Whelan TA. Modifiable partner factors associated with perinatal depression and anxiety: a systematic review and meta-analysis. *J Affect Disord*. 2015;178:165-180. doi:10.1016/j.jad.2015.02.023

35. Bender R, Lange S. Adjusting for multiple testing - when and how? *J Clin Epidemiol*. 2001;54:343-349.

36. De Choudhury M, Counts S, Horvitz EJ, Hoff A. Characterizing and predicting postpartum depression from shared facebook data. In: *Proceedings of the 17th ACM Conference on Computer Supported Cooperative Work & Social Computing*. ; 2014:626-638.

37. Addis ME. Gender and depression in men. *Clin Psychol Sci Pract*. 2008;15(3):153-168. doi:10.1111/j.1468-2850.2008.00125.x

38. Gilbert R, Logan S, Moyer VA, Elliott EJ. Assessing diagnostic and screening tests: Part 1.

Concepts. *West J Med*. 2001;174(6):405-409. doi:10.1136/ewjm.174.6.405

39. Golubnitschaja O, Kinkorova J, Costigliola V. Predictive, preventive and personalised medicine as the hardcore of 'Horizon 2020': EPMA position paper. *EPMA J*. 2014;5(1):6.

40. Teague S, Youssef GJ, Macdonald JA, et al. Retention strategies in longitudinal cohort studies: a systematic review and meta-analysis. *BMC Medical Research Methodology*. 2018;18(1):151. doi:10.1186/s12874-018-0586-7

41. Saha K, Jagannath K, Mark G, et al. Social Media as a Passive Sensor in Longitudinal Studies of Human Behavior and Wellbeing. *Ext Abstr 2019 CHI Conf Hum Factors Comput Syst - CHI '19*. 2019:1-8. doi:10.1145/3290607.3299065

42. De Choudhury M, Kiciman E. Integrating Artificial and Human Intelligence in Complex, Sensitive Problem Domains: Experiences from Mental Health. *AI Mag*. 2018;39(3):69-80. doi:10.1609/aimag.v39i3.2815

43. Ammari T, Schoenebeck S, Romero DM. Self-declared throwaway accounts on Reddit: How platform affordances and shared norms enable parenting disclosure and support. In: *Proceedings of the ACM Conference on Computer Supported Cooperative Work & Social Computing*. ; 2019:Article 135. doi:10.1145/3359237

44. Tsugawa S, Kikuchi Y, Kishino F, Nakajima K, Itoh Y, Ohsaki H. Recognizing depression from twitter activity. *Conf Hum Factors Comput Syst - Proc*. 2015;2015-April:3187-3196. doi:10.1145/2702123.2702280

45. De Choudhury M, Sharma SS, Logar T, Eekhout W, Nielsen RC. Gender and cross-cultural differences in social media disclosures of mental illness. *Proc ACM Conf Comput Support Coop Work CSCW*. 2017;(August 2018):353-369. doi:10.1145/2998181.2998220

46. Cheng Q, Li TM, Kwok C-L, Zhu T, Yip PS. Assessing Suicide Risk and Emotional Distress in Chinese Social Media: A Text Mining and Machine Learning Study. *J Med Internet Res*. 2017;19(7):e243. doi:10.2196/jmir.7276

47. Guntuku SC, Yaden DB, Kern ML, Ungar LH, Eichstaedt JC. Detecting depression and mental illness on social media: an integrative review. *Curr Opin Behav Sci*. 2017;18:43-49. doi:10.1016/j.cobeha.2017.07.005

48. Ernala SK, Birnbaum ML, Candan KA, et al. Methodological gaps in predicting mental health states from social media: Triangulating diagnostic signals. *Conf Hum Factors Comput Syst - Proc*. 2019. doi:10.1145/3290605.3300364

SUPPLEMENTARY FILE 1

Table 1: Variables included in final model after feature selection

Category	Feature
Behaviour	Total number of question submissions
	Total comment score
	Total number of question comments
	Mean number of submissions
	Mean submssion score
	Mean number of submission questions
	Mean number of comments
	Mean comment score
	Mean word count (WC)
	Slope of word count (WC)
Discussion Topic	Total number of posts in discussion topic "*Gender*"
	Total number of posts in discussion topic "*Lifestyle*"
	Total number of posts in discussion topic "*Other*"
	Total number of posts in discussion topic "*Technology*"
	Mean number of posts in discussion topic "*Other*"
	Mean number of posts in discussion topic "*Technology*"
	Slope of posts in discussion topic "*Gender*"
	Slope of posts in discussion topic "*Images*"
	Slope of posts in discussion topic "*Other*"
Linguistic Style	Total number of third person singular pronouns (e.g. she, her, him)
	Total number of third person plural pronouns (e.g. they, their)
	Total number of auxillary verbs (e.g. am, will, have)
	Total number of common adverbs (e.g. very, really)
	Total number of interrogatives (e.g. how, when, what)
	Total number of nonfluencies (e.g. er, hm, umm)
	Mean number of function words (e.g. it, to, no, very)
	Mean number of impersonal pronouns (e.g. it, it's, those)
	Mean number of prepositions (e.g. to, with, above)
	Mean number of auxillary verbs (e.g. am, will, have)
	Mean number of common adverbs (e.g. very, really)
	Mean number of common verbs (e.g. eat, come, carry)
	Mean number of interrogatives (e.g. how, when, what)
	Mean number of nonfluencies (e.g. er, hm, umm)

	Mean number of fillers (e.g. Imean, youknow)
	Slope of of function words (e.g. it, to, no, very)
	Slope of pronouns (e.g. I, them, itself)
	Slope of of third person singular pronouns (e.g. she, her, him)
	Slope of third person plural pronouns (e.g. they, their)
	Slope of impersonal pronouns (e.g. it, it's, those)
	Slope of prepositions (e.g. to, with, above)
	Slope of auxillary verbs (e.g. am, will, have)
	Slope of common adverbs (e.g. very, really)
	Slope of comparison words (e.g. greater, best, after)
	Slope of interrogatives (e.g. how, when, what)
	Slope of quantifiers (e.g. few, many, much)
	Slope of informal language (e.g. swear words, netspeak, filler)
	Slope of swear words (e.g. fuck, damn, shit)
	Slope of "netspeak" (e.g. btw, lol, thx)
	Slope of assent words (e.g. agree, OK, yes)
Emotion	Total number of negative emotion words (e.g. hurt, ugly, nasty)
	Total number of anxiety words (e.g. worried, fearful)
	Total number of anger words (e.g. hate, kill, annoyed)
	Total number of sad words (e.g. crying, grief, sad)
	Mean number of negative emotion words (e.g. hurt, ugly, nasty)
	Mean number of anger words (e.g. hate, kill, annoyed)
	Mean number of sad words (e.g. crying, grief, sad)
	Slope of positive emotion words (e.g. love, nice, sweet)
	Slope of negative emotion words (e.g. hurt, ugly, nasty)
	Slope of anger words (e.g. hate, kill, annoyed)