ORIGINAL RESEARCH ARTICLE

Pharmacoepidemiologic Evaluation of Birth Defects from Health-Related Postings in Social Media During Pregnancy

Su Golder[1] · Stephanie Chiuve[3] · Davy Weissenbacher[2] · Ari Klein[2] · Karen O'Connor[2] · Martin Bland[1] ·
Murray Malin[3] · Mondira Bhattacharya[3] · Linda J. Scarazzini[3] · Graciela Gonzalez-Hernandez[2]

Abstract

Introduction Adverse effects of medications taken during pregnancy are traditionally studied through post-marketing pregnancy registries, which have limitations. Social media data may be an alternative data source for pregnancy surveillance studies.

Objective The objective of this study was to assess the feasibility of using social media data as an alternative source for pregnancy surveillance for regulatory decision making.

Methods We created an automated method to identify Twitter accounts of pregnant women. We identified 196 pregnant women with a mention of a birth defect in relation to their baby and 196 without a mention of a birth defect in relation to their baby. We extracted information on pregnancy and maternal demographics, medication intake and timing, and birth defects.

Results Although often incomplete, we extracted data for the majority of the pregnancies. Among women that reported birth defects, 35% reported taking one or more medications during pregnancy compared with 17% of controls. After accounting for age, race, and place of residence, a higher medication intake was observed in women who reported birth defects. The rate of birth defects in the pregnancy cohort was lower (0.44%) compared with the rate in the general population (3%).

Conclusions Twitter data capture information on medication intake and birth defects; however, the information obtained cannot replace pregnancy registries at this time. Development of improved methods to automatically extract and annotate social media data may increase their value to support regulatory decision making regarding pregnancy outcomes in women using medications during their pregnancies.

Key Points

Social media data can provide information on medication intake and birth defects; however, the information obtained cannot replace pregnancy registries at this time. At present, these data are incomplete but may still be useful to supplement pregnancy registry data.

Future research is necessary to refine efforts and uses of social media data to support regulatory decision making regarding pregnancy outcomes with recently approved drugs used in women of child-bearing age.

✉ Su Golder
su.golder@york.ac.uk

1 Department of Health Sciences, University of York, Heslington, York YO10 5DD, UK

2 Department of Biostatistics, Epidemiology and Informatics, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA, USA

3 AbbVie, North Chicago, IL, USA

△ Adis

1 Introduction

New pharmaceutical products undergo rigorous testing prior to approval. However, data on safety during pregnancy are sparse, particularly for new products [1], as clinical trials often exclude women who are pregnant or breastfeeding because of ethical implications. Up to 80% of pregnant women take at least one prescribed or over-the-counter (OTC) medication during their pregnancy [2]. Further, many pregnancies are unplanned, for example in USA, almost half of pregnancies are unplanned [3], thus unintended fetal exposure to medications during critical periods of development is likely to occur [4]. Therefore, evidence of safety during pregnancy relies on other sources of data, such as observational studies in the post-market setting.

The most common type of pregnancy surveillance studies are traditional pregnancy registries [5], which have been the standard for post-marketing pregnancy surveillance. There are many known disadvantages with registries, including a lack of an appropriate comparator group to estimate background rates, selective loss to follow-up, and low rates of recruitment [6–10]. The routine use of ultrasounds and early prenatal screening challenges true prospective enrollment into traditional registries and may bias results of the registry [9]. Even with efforts made to enroll women as early as possible in the pregnancy (seventh or eighth week of gestation) [1], any adverse drug effects during early pregnancy may be missed [1]. Finally, the frequent inability to have significant statistical associations within pregnancy registries is primarily owing to poor patient enrollment and the few birth defects recorded in these observational studies. Hence, although pregnancy registries have adequate statistical power to detect signals of major-risk teratogenicity (birth defect rate of 25%), they are not powered to detect signals of moderate-risk teratogens [11, 12] or specific birth defects [9]. These limitations, together with the knowledge that one data source is unlikely to be sufficient to provide enough information on potentially rare outcomes, have led researchers and regulatory agencies to identify supplementary sources of data for evaluating the safety of medicines in pregnancy.

Alternative sources for pregnancy surveillance include population-based surveillance registers, electronic healthcare records, and administrative claims databases, and studies within these databases have provided key evidence of drug exposures, pregnancy outcomes, and birth defects [13–20]. However, these sources lack data on OTC medicines and lifestyle factors and prescription fills are used as a surrogate to medication intake. Further, population-based surveillance registers with linkage capabilities between the mother and baby are not available in USA.

Therefore industry-sponsored voluntary registries focused on single drugs or group of drugs associated with a disease (e.g., human immunodeficiency virus, epilepsy) remain the primary source of pregnancy safety surveillance [21, 22].

Social media are another potential emerging source of data for use in pregnancy surveillance. Social media data include information on lifestyle factors collected in women prior to pregnancy and early in the first trimester, when the risk of congenital abnormalities is highest [23]. Other advantages could include prospective data collection in real time and throughout pregnancy and the capture of information on OTC medicines and lifestyle factors, such as smoking and alcohol usage that may be associated with deleterious pregnancy outcomes. In unverified claims, it is postulated that the power of social media to link rare, strikingly unanticipated fetal abnormalities as seen during the "thalidomide storm" to drug usage would have taken only 5–7 days [24].

We selected Twitter for this pilot study as Twitter is a very popular social media source, is publically available, and our prior work has assessed the feasibility of identifying pregnant women who actively use Twitter [25]. For this study, we proposed that publicly available tweets throughout the full timeline of a pregnancy could be annotated, and potential useful information on drug utilization and birth defects obtained. We hypothesized that much of the data routinely collected in registries, such as basic demographics, medicine intake, and birth defects, could be obtained from the social media posts throughout the full term of a pregnancy, and that this annotation could be performed automatically in the future. With data from these timelines, we assessed the feasibility of constructing a nested case-control study within a cohort of pregnant women to quantify the association between pregnancy-related exposure and birth defects.

2 Methods

The methodology followed for this study needed to address the problems of identifying women having a baby with a birth defect using primarily automated methods. Given the rare occurrence of birth defects, case-control studies nested within large populations are the preferred approach for the evaluation of specific pregnancy outcomes [6]. Thus, to identify the case and control groups from social media, we first needed to identify pregnant women amongst the millions of Twitter users. Our initial work was focused on this, detecting users to add to our pregnancy database via a single tweet announcement [25]. Our automatic classification system achieves an F_1-score of 0.88 for identifying pregnancies. The F_1-score is computed as $2 \times (Precision \times Recall)/(Precision + Recall)$, where precision is *True Positives/(True Positives + False Positives)*,
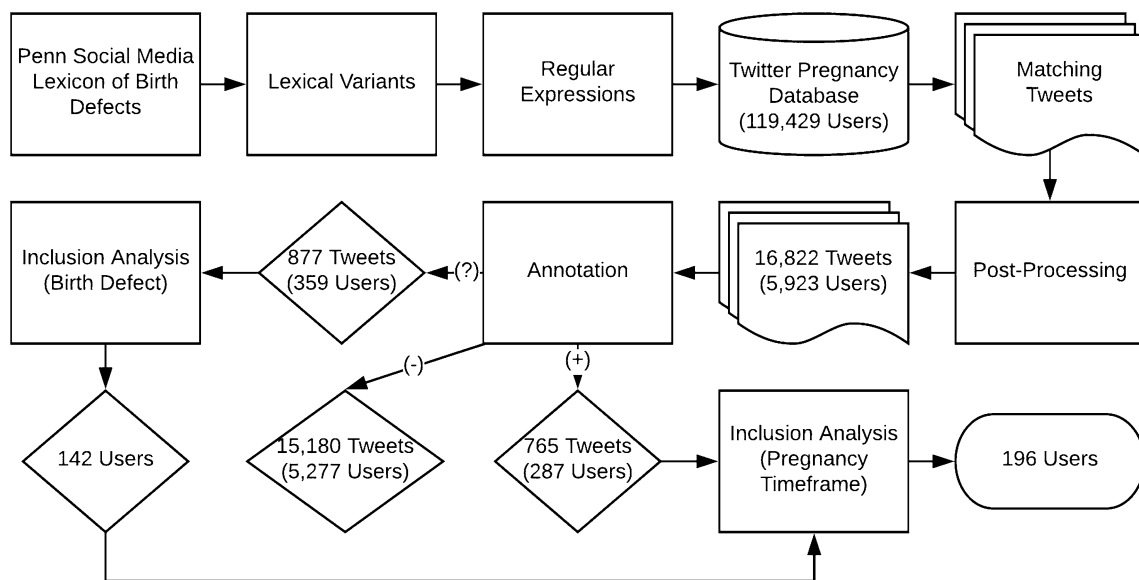
Fig. 1 Workflow of tweet collection, tweet annotation, and timeline analysis for selecting the birth defects (case) cohort

and recall is *True Positives/(True Positives + False Negatives)*. Once pregnant users are identified, all of their publicly available tweets (their "timeline") were collected. A total of 112,429 users were identified and their timelines collected. A method for estimating the number of timelines that encompasses the user's pregnancy was developed [26], resulting in a total of 44,825 timelines in our database.

2.1 Selection of the Cohorts

A birth defects cohort (cases) was created by retrieving and annotating tweets from the pregnancy database that mention birth defects. This method, which we summarize in the remainder of this sub-section, is described in further detail in another publication [26]. As Fig. 1 illustrates, we manually compiled a lexicon of approximately 650 terms referring to birth defects (Penn Social Media Lexicon of Birth Defects), based on published reports, guidelines, and the Unified Medical Language System [27–31], and semi-automatically generated lexical variants of these terms (e.g., misspellings). To retrieve tweets containing (variants of) the terms, we implemented hand-crafted regular expressions in a series of database queries. We post-processed the retrieved tweets by removing those containing user names and URLs matched by the regular expressions. With this retrieval method, a total of 16,822 tweets (posted by 5923 users) were collected, with a recall of 0.95. The tweets were annotated by two annotators. We developed annotation guidelines to distinguish three classes of tweets, summarized as follows:

Defect (+)	The tweet refers to a person who has a birth defect and identifies that person as the Twitter user's child
Possible Defect (?)	The tweet is ambiguous about whether a person referred to has a birth defect and/or is the Twitter user's child
Non-defect (−)	The tweet does not indicate that a person referred to has or may have a birth defect and is or may be the Twitter user's child

The annotators' inter-annotator agreement (Cohen's kappa) was high ($\kappa = 0.79$). In total, 765 (4.55%) tweets were annotated as "defect," 877 (5.21%) tweets were annotated as "possible defect," and 15,180 (90.24%) tweets were annotated as "non-defect." The annotations directed us to the timelines of the users who posted them, for an inclusion/exclusion analysis to determine a final cohort. Users were excluded from the cohort if we could not determine if they were the parent of a child with a birth defect, or if there were no tweets available during the pregnancy with a birth defect outcome. First, we analyzed the timelines of the 359 users who posted a "possible defect" tweet (without also posting a "defect" tweet), and determined that 142 (39.55%) of them were indeed the parent of a child with a birth defect. Then, we analyzed the timelines of these 142 users and the 287 users who posted a "defect" tweet, and determined that 196 (45.69%) of the 429 timelines encompass tweets from the timeframe of the pregnancy with a birth defect outcome. Thus, we identified 196 users for our birth defects (case) cohort. For this study, the timelines of the 196 women reporting a birth defect (cases) were matched on timing of

pregnancy to timelines of 196 women not reporting any birth defects (controls) in the pregnancy database.

2.2 Data Preparation

All tweets mentioning birth defects were automatically identified using the lexical approach presented in [26]. For this project, we retrieved the timelines of the users corresponding to these tweets and tagged recognizable medication names to facilitate the manual annotation process of the timelines. A set of 37 drug names, including variants and misspellings, was already annotated in our timelines and manually classified into *intake*, *possible intake*, or *no intake* categories. For greater coverage of medications, we extended this initial set of drugs with the list of drug names published in the Drugs@FDA database.[1] We added lexical variants (possible misspellings) of these drug names and tagged the names in the timelines. All drug mentions found in the tweets during this last process were then automatically classified as *intake*, *possible intake*, or *no intake* using our in-house classifier [32]; this pre-annotating step speeds up manual curation. In our past work [32], inter-annotator agreement (Cohen's kappa) for manually identifying medication intake was very high ($\kappa = 0.88$). Finally, all mentions of gestational ages were automatically pre-annotated and tagged in the timelines [33].

2.2.1 Annotation of Exposures of Interest

To analyze the data for the cases and controls, we first needed to manually annotate the timelines for exposures of interest whenever we did not have any automatic method of doing so, and to corroborate any information tagged automatically. Exposures of interest were; maternal age, due date, place of residence, race/ethnicity, medicine intake at first, second, and third trimester, and birth defects. We created an annotation guideline with examples, and selected the General Architecture for Text Engineering environment for annotation [34].

We annotated all tweets in *the timeline of the pregnancy,* defined as the time of pregnancy plus 1 month before and 1 month after. Within this timeframe, we annotated all mentions of *gestational ages*, any indications of the *due date* of delivery, the *pregnancy outcome*, and the *date of birth* of the child. We also annotated each tweet mentioning a drug name listed in the *Drugs@FDA* database and if the drug was taken (or possibly taken) by trimester of the pregnancy. We annotated all mentions of birth defects and then classified them under their corresponding *Medical Dictionary for*

Regulatory Activities categories. Annotation guidelines are provided as a supplement.

Maternal age was often given in reference to a birthday such as "I'm 24 on Friday" or 'Only 2 hours until I'm 21 and legal!' Others mentioned their age in passing "I'm 22 but look 26" or "Gosh you would think I would know that at 20." Where only approximations of age were given (e.g., women indicated that they were in their 20s or 30s), we categorized these as missing data.

The country of residence of the woman was often present in the profile information (e.g., "Proud Colombian," "Texas," or "Bangor, Wales") or stated in a post. Race was also sometimes explicitly stated: "Just because I'm Hispanic …" or "I'm not African-American I'm black American.".

Medications were categorized based on the available evidence of risks associated with taking particular medicines while pregnant as per the Australian categorization system.[2] We selected this categorization system as there is no language barrier, it has greater granularity of classifications (with seven categories A, B1, B2, B3, C, D, X), is easy to use, and is up to date. Some comments were not possible to classify as there was insufficient detail to identify the medication, such as "My pain meds aren't working anymore," or "Got to take antibiotics for …". Medications were grouped into 'probably safe' or 'potentially risky' to help facilitate the analysis and compare with previous studies. [35] The 'probably safe' category consisted of A, B1, and B2 classifications and the 'potentially risky' category consisted of B3, C, D, and X.

Although in most instances the medications were named in the tweets, we made a concerted effort not to publish the individual drug names in this article. This was because the study was a feasibility study to test the methodologies using social media. Given the exploratory nature of these methods, we chose not to study individual drug products and raise concern over spurious safety signals without further evidence of causality. To gather additional information on drug classes would require additional data mining and natural language processing information that was not collected initially and is beyond the scope of this project.

Most of the comments on medication intake refer to when the medication was consumed either directly or indirectly: "4 mg of X four hours ago and still got a headache," "taken X and now off to bed." We were able to ascertain the timing of the medication intake for the users by the date of the post and then assess whether the intake was from the first, second, or third trimester. Many women gave the actual due date for their baby or provided information from which the due date could be calculated "I am due on the 24th of February" or

---

Table 1 Birth defects by the *Medical Dictionary for Regulatory Activities* (MedDRA)

MedDRA sub-class of congenital, familial, and genetic disorders	Frequency
Musculoskeletal and connective tissue disorders congenital	63
Cardiac and vascular disorders congenital	46
Gastrointestinal tract disorders congenital	45
Chromosomal abnormalities and abnormal gene carriers	21
Neurological disorders congenital	13
Renal and urinary tract disorders congenital	5
Eye disorders congenital	3
Respiratory disorders congenital	3
Skin and subcutaneous tissue disorders congenital	2
Blood and lymphatic system disorders congenital	2
Hepatobiliary disorders congenital	1
Reproductive tract and breast disorders congenital	1
Ear and labyrinth disorders congenital	1

The sum of the frequencies is slightly greater than the number of cases; this is because, in some of the individual cases, the child had multiple birth defects and the defects belonged to different sub-classes

"I'm due a week today." There were many references to the length of term of the pregnancy (either by how far into the pregnancy they were or by how long the pregnancy had left). For instance, "I'm 24 weeks today" or "6 more weeks to my due date." From gestational age annotations, the annotator could calculate an estimated pregnancy conception date using an Internet pregnancy calculator.[3] From this information, exposure to medications could be categorised as in the first, second, or third trimester.

2.3 Statistical Methods

Using the cc command in Stata, we estimated the odds ratios for each risk factor. The confidence intervals were estimated by the exact method. For the type of medication, we calculated an overall p value using a chi-squared test for a contingency table, as these are not independent variables but different categories of the same variable. A p value < 0.05 was considered statistically significant.

To check whether matching was informative, the analysis for any medication use was also carried out by logistic regression, both ignoring the matching and allowing for matching using robust standard errors. Results were not appreciably altered and, therefore, we conducted an analysis without matching to minimize the impact of missing data.

To check whether other available risk factors (age, ethnicity, country of residence) could explain the relationship between birth defects and medication, logistic regression was used. We excluded women for whom any of these

variables were missing and for ethnicity and residence categories with fewer than ten women, we used the "other" category.

3 Results

The mean number of posts per pregnancy timeline in the cases was 2903, varying from 70 to 15,271 posts. The average number of posts per woman in the control group was lower at 2582 (range 19–9142). For comparison, in the entire database of 112,429 timelines, the mean number of posts per person was 3850 (range: 2–80,023 posts). Annotations of each of the 196 cases and 196 control pregnancy timeline took an average of 2 h.

The rate of birth defects in our cohort of pregnant women (cases) was 0.44%. We calculated this rate by taking the ratio between the number of pregnant women reporting a birth defect and the estimated total number of timelines in our database in which we had access to the user's tweets during pregnancy (i.e., 196/44,825). Examples of birth defects included cleft lip, club foot, congenital heart defects, and Down's syndrome. The frequencies of the consolidated birth defects reported in our cases are presented in Table 1. The categories used were based on *Medical Dictionary for Regulatory Activities* sub-classes.

3.1 Characteristics of Women

Women who gave birth to a baby with a birth defect had a different demographic profile than women who gave birth to baby without a birth defect. Cases were older, more likely to be Caucasian, and less likely to live in USA; cases were also more likely to have missing information on race and

---

[3] Available at https://www.babymed.com/tools/retro-conception-birth day.

Table 2 Characteristics of the cases and controls among women who gave birth

Characteristics	Cases ($n = 196$)	Controls ($n = 196$)	p value, difference[a]
Age, years			
Median age (IQR)	23 (20–28)	21 (19–23)	0.0001
Mean age (range)	25 (17–42)	22 (16–37)	< 0.0001
Women < 30, %	68	66	0.004
Women < 35, %	80	70	0.04
Missing data on age, %	14	28	0.0008
Race/ethnicity, %			
Caucasian	61	52	< 0.001
Black	11	26	
Hispanic	6	11	
Asian	2	3	
Other	2	3	
Missing data on race	16	6	
Place of residence, %			
USA	66	77	0.04
UK	16	8	
Canada	4	3	
Other	2	3	
Missing data on place of residence	6	9	

IQR interquartile range

[a]p values were estimated using the chi-squared test
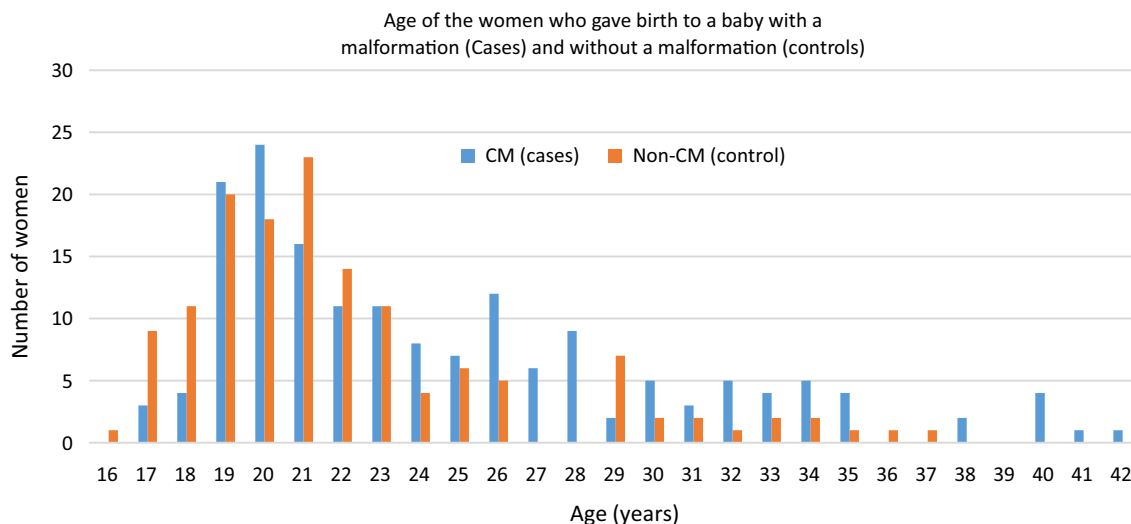
Fig. 2 Age of the women who gave birth to a baby with a birth defect (cases) and without a birth defect (controls). *CM* women who gave birth to a baby with a malformation

less likely to have missing data on age. The distributions of age in both cases and controls are presented in Table 2 and Fig. 2.

3.2 Timing and Type of Medication Intake

Cases reported taking some form of medication during pregnancy more frequently than the controls (35% vs. 17%) (Table 3). Many women, particularly in the cases, mentioned more than one medication intake or taking the same medication on more than one occasion (Table 4).

In the first and third trimesters, the number of women taking medication among the cases was significantly higher than in the controls [odds ratio (OR) = 3.59 (1.44–10.13); $p = 0.002$ and OR = 2.22 (1.23–4.08); $p = 0.004$, respectively]. In the second trimester, although a higher number of

Table 3 Medication intake, timing, and type in the cases and controls

	Cases, % (n/N)	Controls, % (n/N)	p value[a]
Medication use			
Any medication use during pregnancy	35 (68/196)	17 (34/196)	0.0001
Timing of medication intake among women taking medications[b]			
Any medication use during first trimester[d]	42 (23/55)	23 (7/31)	0.1
Any medication use during second trimester[d]	34 (22/65)	38 (12/32)	0.7
Any medication use during third trimester	63 (43/68)	65 (22/34)	0.9
Type of medication intake among women taking medications[b]			
'Probably safe' medications only[c]	62 (42/68)	65 (22/34)	0.9
At least one 'potentially risky' medication[c]	21 (14/68)	18 (6/34)	
At least one unclassified medication[c]	18 (12/68)	18 (6/34)	

The 'probably safe' category consisted of A, B1, and B2 classifications and the 'potentially risky' category consisted of B3, C, D, and X as per the Australian categorization system (https://www.tga.gov.au/prescribing-medicines-pregnancy-database)

[a]p values were estimated using the chi-squared test

[b]'Among women taking medications' means the denominator used was only those women who reported taking any medication as opposed to the whole group of women

[c]Multiple medications were taken by some women and some medications were taken more than once

[d]For some women, data from posts were missing in the first and second trimester

Table 4 Percentage of instances of intake of 'probably safe', 'potentially risky', and 'unclassified' medications

	Instances in cases, % (n/N)	Instances in controls, % (n/N)	p value[a]
First trimester, N	27	8	
'Probably safe' medication	67 (18/27)	63 (5/8)	0.70
'Potentially risky' medication	22 (6/27)	13 (1/8)	
'Unclassified' medication	11 (3/27)	25 (2/8)	
Second trimester, N	39	14	
'Probably safe' medication	61 (24/39)	64 (9/14)	0.75
'Potentially risky' medication	15 (6/39)	21 (3/14)	
'Unclassified' medication	23 (9/39)	14 (2/14)	
Third trimester, N	60	26	
'Probably safe' medication	73 (44/60)	77 (20/26)	0.93
'Potentially risky' medication	15 (9/60)	15 (4/26)	
'Unclassified' medication	12 (7/60)	8 (2/26)	
Total pregnancy, N	126	48	
'Probably safe' medication	68 (86/126)	71 (34/48)	0.97
'Potentially risky' medication	17 (21/126)	17 (8/48)	
'Unclassified' medication	15 (19/126)	13 (6/48)	

The 'probably safe' category consisted of A, B1, and B2 classifications and the 'potentially risky' category consisted of B3, C, D, and X as per the Australian categorization system (https://www.tga.gov.au/prescribing-medicines-pregnancy-database)

[a]Fisher's exact t test

women among the cases took medications than the controls (22 vs. 12), this was not significantly different [OR = 1.94 (0.89–4.43); p = 0.07].

When the analysis was restricted to women who reported taking medications during pregnancy, the pattern in timing of intake among the cases and controls was similar (Table 3). This is demonstrated by the fact that there is no statistically significant difference *between the women taking any medication* in the timing of their intake between the cases and controls (p = 0.1, p = 0.7, p = 0.9 for the first, second, and third trimester respectively) (Table 3).

There were 53 different medications reported as taken in the timelines in the cases and 24 different medications mentioned in the control timelines. The number of women taking 'probably safe medications only', 'at least one potentially risky medication', or 'at least one unclassified medication' was higher in the cases than in the controls (42/196, 21% vs. 22/196, 11%, 14/196, 14% vs. 6/196, 3%, 12/196, 6% vs. 6/196, 3%, respectively) (Table 3). If we limit our analysis to only those women who reported taking at least one medication, we find that the pattern of intake by type of medication is very similar in the cases and controls (62% vs. 65%, 21% vs. 18%, 18% vs. 18%) [p = 0.9] (Table 3).

3.3 Predictors of Birth Defects

Using logistic regression, medication use was associated with a greater risk of birth defects [OR = 2.53; p < 0.001, 95% confidence interval (CI) 1.58–4.06]. This result was not appreciably altered after adjusting for age, ethnicity, and country of residence. In multivariable models, the

Table 5 Odd ratios (95% confidence intervals) [ORs (95% CIs)] for birth defects by various demographic and lifestyle factors

Variable	OR (95% CI) [logistic regression estimates]			
	Univariable, unadjusted	*p* value	Multivariable, adjusted for other variables	*p* value
Age (per year)	1.10 (1.05–1.15)	< 0.001	1.09 (1.03–1.15)	0.002
Medication use				
Yes	2.53 (1.58–4.06)	< 0.001	2.34 (1.24–4.44)	0.004
No	1.0 (ref)		1.0 (ref)	
Ethnicity				
Caucasian	1.0 (ref)	< 0.001	1.0 (ref)	0.008
Black	0.37 (0.21–0.65)		0.40 (0.21–0.79)	
Hispanic	0.57 (0.27–1.17)		0.86 (0.36–2.04)	
Asian	0.68 (0.18–2.60)		0.83 (0.25–1.55)	
Other	0.68 (0.18–2.60)		0.80 (0.19–3.48)	
Missing	2.27 (1.11–4.62)		3.42 (1.25–9.40)	
Place of residence				
USA	1.0 (ref)	0.01	1.0 (ref)	0.3
UK	2.28 (1.19–4.36)		1.97 (0.85–4.57)	
Canada	1.89 (0.60–5.90)		1.06 (0.22–5.03)	
Other	0.20 (0.02–1.65)		0.24 (0.01–4.43)	
Missing	1.73 (0.90–3.35)		1.69 (0.69–4.11)	

ref reference

association between any medication use and the risk of birth defects was slightly reduced (OR = 2.34; $p = 0.004$, 95% CI 1.24–4.44), but it remained highly significant (Table 5).

Conducting a one-factor analysis for age, ethnicity, and residence, we found that older women were more likely to report birth defects [age (per year): OR = 1.10, $p < 0.001$, 95% CI 1.05–1.15] (Table 5). However, when ethnicity and residence were included in the model as categorical factors, ethnicity was statistically significant ($p = 0.008$), country of residence was not ($p = 0.3$). For both categorical variables, missing was included as a category, and caution must be taken in interpreting these results.

4 Discussion

We have demonstrated that there is a large amount of data publicly available on social media, specifically Twitter, from women during their pregnancy and on their pregnancy outcomes, with many women posting on a daily basis. From these data, we created a prospective timeline for women posting on social media regarding their pregnancies. We were also able to extract information on birth defects, lifestyle factors, and medication intake, including the frequency, timing, and type of medication use before and during the gestational period. However, the main results of the pilot study demonstrated a rate of malformations lower (0.44%) than the rate reported in the general population (3%), highlighting incompleteness and bias in social media data with respect to sensitive medical information such as birth defects.

Our analytic approach to social media data included a nested case-control study comparing exposures among women who gave birth to babies with a birth defect to women whose baby did not have a birth defect. We found that women who gave birth to babies with a birth defect were more likely to be older, Caucasian, and live outside of USA. Even after accounting for age, race, and place of residence between cases and controls, a higher medication intake was observed in pregnancies that reported birth defects. However, women who gave birth to babies with a birth defect also had a higher rate of missing data, limiting the causal inferences that can be made from this analysis. As automated methods for annotation of key demographic, medical and social data are further refined and validated, a nested case-control study design will be the ideal study design to assess pregnancy outcomes from social media data sources because of the rarity of birth defects in the population [6].

There are a number of potential benefits of social media data as an alternative to pregnancy registries. First, even if the women may be identified later in their pregnancies, data are collected prospectively, therefore reducing or eliminating recall bias. Other advantages are the potential availability of data on OTC medications, illicit drugs, and lifestyle factors such as smoking and alcohol that are not captured during routine healthcare encounters and other secondary data sources. Another benefit of social media

data is that a comparator group of unexposed pregnant women can be ascertained, which is often lacking in traditional registries. Additionally, although not the focus of our investigation, social media posts contained many adverse pregnancy outcomes, such as early pregnancy loss, low birth weight, and premature delivery, which are not the primary outcomes of interest in pregnancy registries.

Conversely, there are several drawbacks of social media data. First, there are potential differences in key factors associated with birth defects when compared with the general population [36]. For example, the mean age of the cases and controls in this study was approximately 7 and 9 years younger than the general population, respectively, although this may be a reflection of the large number of women classified as having a missing age [37, 38]. Other factors, such as the education levels and social class of social media users, may differ from the wider population. Additionally, the proportion of women reporting at least one medication use is low in both our cases and controls compared with other studies [2]. Finally, the rate of birth defects in the social media population was lower than the rate in the general US population [39, 40].

Reasons for the underestimation in the current study include the incompleteness or underreporting of key information as a result of multiple factors, such as the fact that women may be less likely to report high-risk behaviors and women who are aware that their babies may have a birth defect may be less likely to discuss this information on social media. Additionally, the Natural Language Processing method used to identify birth defects might not be able to capture all such mentions and requires further development. Many women did not allude to the birth defects in much detail or with as much frequency as would be expected given the detail in their other posts while pregnant. Some women also played down any birth defect, posting remarks such as "it's no big deal but …".

Recent reviews of traditional pregnancy registries conducted by the US Food and Drug Administration have identified key challenges in the recruitment of patients including the reduced likelihood of women to continue to use drugs that may be associated with birth defects [41] and the widespread use of early prenatal screening [8]. Social media have the potential to identify women for recruitment into traditional registries even prior to conception, women who are exposed and unexposed to the drug of interest, and to reduce the recall bias associated with key lifestyle and medical factors contributing to birth defects. This ability to target and recruit women from a larger pool would allow for the assessment of birth defects with greater statistical power, and the availability of non-exposed women provides greater clinical relevance to these statistical findings. The ethical issues around this active recruitment method need careful consideration [42]. It is anticipated that ethical approval and informed consent will be required to collect information in this manner and for its use for research purposes. Guidelines to help assist researchers to consider the ethical issues for the many different approaches to using social media in research are available and continue to be developed [42].

The information from social media could also be used to inform public health and health promotion campaigns. Some of the medications identified among this study cohort are known to cause problems during pregnancy [43] and many have safer alternatives. The patterns of medication intake could be used to prioritize which medications should be highlighted as potentially unsafe during pregnancy in public health messages. For instance, the risks associated with the use of ibuprofen during pregnancy may not be understood by women. While data are mixed, non-steroidal anti-inflammatory drugs such as ibuprofen have been linked to an increased risk of spontaneous abortion and congenital malformations when taken in the first trimester [43–45] and linked to renal impairment and cardiopulmonary abnormalities in the neonate when taken later in pregnancy [46]. Additionally, there have been reports on an increased risk of postpartum hemorrhage for women exposed to non-steroidal anti-inflammatory drugs [44].

4.1 Limitations

Automatic language processing methods utilized in this study enabled the selection of pregnant women from social media. These methods also facilitated the identification of concepts of interest (birth defects, medication intake, and pregnancy timeframe) and greatly reduced the annotation effort. However, the manual annotation effort for the identification of birth defects, which was the primary focus of this study, still required 800 h to annotate over 100,000 tweets, which limited the ability to include a greater number of controls for each pregnant case and to extract additional valuable information available within these tweets. The amount and detail of information disclosed in the pregnancy timelines was considerable and sometimes overwhelming. The amount of information varies from individual to individual, with some twitter users disclosing many personal thoughts to others who limit the personal information they choose to post. Twitter has recently increased the number of characters allowed on each post from 140 characters to 280 characters. This may increase the level of detail posted and lead to less ambiguous posts and improve data clarity, while also increasing the annotation burden. Additional methodologic challenges included the inability to match cases and controls by maternal age, despite age being the biggest risk factor for birth defects. Not all users had their age in their profile information or posted in their tweets. Further automatic language processing advances are warranted to improve these methods and to develop new methods to automatically extract other

relevant data from social media timelines (such as pregnancy outcomes, age, place of residence, race, and later substance use) for rapid safety surveillance. For example, automatic methods to determine age from a timeline would have facilitated a greater than 1:1 match between cases and controls. Additional research to develop automatic methods for detecting birth defects through social media data is needed. With additional inputs and broader algorithms, we may be able to capture additional pregnancies with birth defects in future work.

5 Conclusion

In future research, the study design should ideally incorporate matching of cases and controls by key factors including age, race, geography, gestational timeline, and volume of tweets to have greater certainty on conclusions drawn regarding associations between drug use and outcomes of interest. The specific focus on matching on volume of tweets is to reduce the likelihood that key medical information would be missing and we need to consider how to reduce the chance of false negatives among the control arms. Further, it cannot be assumed that no mention of a birth defect or medication intake indicates that no such event occurred. Therefore, a validation of cases of birth defects identified through pregnancy timelines against diagnoses from medical records would provide additional certainty regarding the specificity and sensitivity for case ascertainment. The associations (positive or negative) derived from social media data should be validated against the association estimated from other data sources, including voluntary registries and claims/electronic healthcare record databases. These validation efforts will be required to use results from studies in social media data sources in submissions to regulatory agencies as an alternative to traditional voluntary registries. Other types of social media (particularly non-microblogging sites) should be investigated as different results may be obtained. Finally, future research is needed to determine in which scenarios social media data may be most informative, including for which drug types, frequency of exposure, and magnitude of association.

While social media data have their limitations, in this pilot effort, we have demonstrated that it is feasible to use Twitter data in assessing medication intake and birth defects; however, the information obtained cannot replace pregnancy registries at this time. With further refinement and validation, social media data could potentially complement other established methods in further characterizing the effects of drugs after introduction to the market, including populations underrepresented or not studied (i.e., pregnant women) in clinical development programs. Development of improved methods to automatically extract and annotate social media data may increase their value in supporting regulatory decision making regarding pregnancy outcomes in women using medications during their pregnancies

Compliance with Ethical Standards

References

1. Chambers C, Cohen L, Koren G. Pregnancy registries: advantages and disadvantages. Ob Gyn News 2012.
2. Lupattelli A, Spigset O, Twigg MJ, Zagorodnikova K, Mardby AC, Moretti ME, et al. Medication use in pregnancy: a cross-sectional, multinational web-based study. BMJ Open. 2014;4(2):e004365. https://doi.org/10.1136/bmjopen-2013-004365.
3. Finer LB, Zolna MR. Declines in unintended pregnancy in the United States, 2008–2011. N Engl J Med. 2016;374(9):843–52. https://doi.org/10.1056/nejmsa1506575.
4. Koren G, Pastuszak A, Ito S. Drugs in pregnancy. N Engl J Med. 1998;338(16):1128–37. https://doi.org/10.1056/nejm199804163381607.
5. Drug Safety Research Unit. Registries. 2018. http://www.dsru.org/studies-for-risk-management/registries/. Accessed 21 Sept 2018.
6. Hernandez-Diaz S, Oberg AS. Are epidemiological approaches suitable to study risk/preventive factors for human birth defects? Curr Epidemiol Rep. 2015;2(1):31–6. https://doi.org/10.1007/s40471-015-0037-5.
7. Krueger WS, Anthony MS, Saltus CW, Margulis AV, Rivero-Ferrer E, Monz B, et al. Evaluating the safety of medication exposures during pregnancy: a case study of study designs and data sources in multiple sclerosis. Drugs Real World Outcomes. 2017;4(3):139–49. https://doi.org/10.1007/s40801-017-0114-9.

8. Bird ST, Gelperin K, Taylor L, Sahin L, Hammad H, Andrade SE, et al. Enrollment and retention in 34 United States pregnancy registries contrasted with the manufacturer's capture of spontaneous reports for exposed pregnancies. Drug Saf. 2018;41(1):87–94. https://doi.org/10.1007/s40264-017-0591-5.

9. Gelperin K, Hammad H, Leishear K, Bird ST, Taylor L, Hampp C, et al. A systematic review of pregnancy exposure registries: examination of protocol-specified pregnancy outcomes, target sample size, and comparator selection. Pharmacoepidemiol Drug Saf. 2017;26(2):208–14. https://doi.org/10.1002/pds.4150.

10. Gliklich RE, Dreyer NA, Leavy MB (eds). Registries for evaluating patient outcomes: a user's guide. Rockville (MD): Agency for Healthcare Research and Quality (US). 2014.

11. Mitchell AA. Systematic identification of drugs that cause birth defects: a new opportunity. N Engl J Med. 2003;349(26):2556–9. https://doi.org/10.1056/nejmsb031395.

12. Charlton RA, Cunnington MC, de Vries CS, Weil JG. Data resources for investigating drug exposure during pregnancy and associated outcomes: the General Practice Research Database (GPRD) as an alternative to pregnancy registries. Drug Saf. 2008;31(1):39–51.

13. Bateman BT, Patorno E, Desai RJ, Seely EW, Mogun H, Dejene S, et al. Angiotensin-converting enzyme inhibitors and the risk of congenital malformations. Obs Gyn. 2017;129(1):174–84. https://doi.org/10.1097/aog.0000000000001775.

14. Petersen I, Sammon CJ, McCrea RL, Osborn DPJ, Evans SJ, Cowen PJ, et al. Risks associated with antipsychotic treatment in pregnancy: comparative cohort studies based on electronic health records. Schizophr Res. 2016;176(2–3):349–56. https://doi.org/10.1016/j.schres.2016.07.023.

15. Huybrechts KF, Hernandez-Diaz S, Avorn J. Antidepressant use in pregnancy and the risk of cardiac defects. N Engl J Med. 2014;371(12):1168–9. https://doi.org/10.1056/nejmc1409203.

16. Baril L, Rosillon D, Willame C, Angelo MG, Zima J, van den Bosch JH, et al. Risk of spontaneous abortion and other pregnancy outcomes in 15–25 year old women exposed to human papillomavirus-16/18 AS04-adjuvanted vaccine in the United Kingdom. Vaccine. 2015;33(48):6884–91. https://doi.org/10.1016/j.Vaccine.2015.07.024.

17. Huybrechts KF, Bateman BT, Palmsten K, Desai RJ, Patorno E, Gopalakrishnan C, et al. Antidepressant use late in pregnancy and risk of persistent pulmonary hypertension of the newborn. JAMA. 2015;313(21):2142–51. https://doi.org/10.1001/jama.2015.5605.

18. Thyagarajan V, Robin Clifford C, Wurst KE, Ephross SA, Seeger JD. Bupropion therapy in pregnancy and the occurrence of cardiovascular malformations in infants. Pharmacoepidemiol Drug Saf. 2012;21(11):1240–2. https://doi.org/10.1002/pds.3271.

19. Green MW, Seeger JD, Peterson C, Bhattacharyya A. Utilization of topiramate during pregnancy and risk of birth defects. Headache. 2012;52(7):1070–84. https://doi.org/10.1111/j.1526-4610.2012.02190.x.

20. Molgaard-Nielsen D, Hviid A. Newer-generation antiepileptic drugs and the risk of major birth defects. JAMA. 2011;305(19):1996–2002. https://doi.org/10.1001/jama.2011.624.

21. Watts DH, Covington DL, Beckerman K, Garcia P, Scheuerle A, Dominguez K, et al. Assessing the risk of birth defects associated with antiretroviral exposure during pregnancy. Am J Obs Gyn. 2004;191(3):985–92. https://doi.org/10.1016/j.ajog.2004.05.061.

22. Hernandez-Diaz S, Smith CR, Shen A, Mittendorf R, Hauser WA, Yerby M, et al. Comparative safety of antiepileptic drugs during pregnancy. Neurology. 2012;78(21):1692–9. https://doi.org/10.1212/wnl.0b013e3182574f39.

23. Charlton R, de Vries C. Systematic overview of data sources for drug safety in pregnancy research. Consultancy EMA/2010/29/CN. prepared for the European Medicines Agency, June 2012 updated for ENCePP June 2016. http://www.encepp.eu/structure/documents/Data_sources_for_medicines_in_pregnancy_research.pdf.

24. Tucker E. How pharmaceuticals can avoid the side effects of social media. MIT Sloan Management Review. 2013. https://sloanreview.mit.edu/article/how-pharmaceuticals-can-avoid-the-side-effects-of-social-media/. Accessed 23 Sept 2018.

25. 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. https://doi.org/10.2196/jmir.8164.

26. 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.

27. Sever LE. Guidelines for conducting birth defects surveillance. National Birth Defects Prevention Network. 2004. www.nbdpn.org/bdsurveillance.html. Accessed 23 Sept 2018.

28. Correa A, Cragan JD, Kucik JE, Alverson CJ, Gilboa SM, Balakrishnan R, et al. Reporting birth defects surveillance data 1968–2003. Birth Defects Res A Clin Mol Teratol. 2007;79(2):65–186. https://doi.org/10.1002/bdra.20350.

29. Fornoff JE, Shen T. Birth defects and other adverse pregnancy outcomes in Illinois 2005–2009: a report on county-specific prevalence. 2013. http://www.dph.illinois.gov/sites/default/files/publications/ers14-03-birth-defects-inillinois-2005-2009-041516.pdf. Accessed 23 Sept 2018.

30. EUROCAT. Guide 1.4: instruction for the registration of congenital anomalies. 2013. http://www.eurocat-network.eu/content/Section%203.3-%2027_Oct2016.pdf. Accessed 23 Sept 2018.

31. US National Library of Medicine. UMLS reference manual. 2009. https://www.ncbi.nlm.nih.gov/books/NBK9676/pdf/Bookshelf_NBK9676.pdf. Accessed 23 Sept 2018.

32. Klein AZ, Sarker A, Rouhizadeh M, O'Connor K, Gonzalez G. Detecting personal medication intake in Twitter: an annotated corpus and baseline classification system. Proceedings of the BioNLP 2017 workshop. Vancouver (BC): ACM Press; 2017. pp. 136–42. http://www.aclweb.org/anthology/W17-2316. Accessed 23 Sept 2018.

33. Rouhizadeh M, Magge A, Klein A, Sarker A, Gonzalez G. A rule-based approach to determining pregnancy timeframe from contextual social media postings. In: Proceedings of the 2018 International Conference on Digital Health—DH '18. New York (NY): ACM Press; 2018. pp. 16–20. http://dl.acm.org/citation.cfm?doid=3194658.3194679. Accessed 23 Sept 2018.

34. Cunningham H, Tablan V, Roberts A, Bontcheva K. Getting more out of biomedical documents with GATE's full lifecycle open source text analytics. PLoS Comput Biol. 2013;9(2):e1002854. https://doi.org/10.1371/journal.pcbi.1002854.

35. Tronnes JN, Lupattelli A, Nordeng H. Safety profile of medication used during pregnancy: results of a multinational European study. Pharmacoepidemiol Drug Saf. 2017;26(7):802–11. https://doi.org/10.1002/pds.4213.

36. Giles EL, Adams JM. Capturing public opinion on public health topics: a comparison of experiences from a systematic review, focus group study, and analysis of online, user-generated content. Front Public Health. 2015;3:200. https://doi.org/10.3389/fpubh.2015.00200.

37. Centers for Disease Control and Prevention's National Center for Health Statistics. Mean age of mothers is on the rise: United States, 2000–2014. NCHS Data Brief No. 232. 2016. https://www.cdc.gov/nchs/products/databriefs/db232.htm. Accessed 21 Sept 2018.

38. Office of National Statistics. Statistical bulletin: births by parents' characteristics in England and Wales. 2016. https://www.ons.gov.uk/peoplepopulationandcommunity/birthsdeathsandmarriages/livebirths/bulletins/birthsbyparentscharacteristicsinenglandandwales/2016. Accessed 21 Sept 2018.

39. Parker SE, Mai CT, Canfield MA, Rickard R, Wang Y, Meyer RE, et al. Updated national birth prevalence estimates for selected birth defects in the United States, 2004–2006. Birth Defects Res A Clin Mol Teratol. 2010;88(12):1008–16. https://doi.org/10.1002/bdra.20735.

40. Hoyert DL, Xu J. Deaths: preliminary data for 2011. National vital statistics reports, vol 61. Hyattsville (MD): National Center for Health Statistics; 2012. p. 1–52. https://www.cdc.gov/nchs/data/nvsr/nvsr61/nvsr61_06.pdf.

41. Illoh OA, Toh S, Andrade SE, Hampp C, Sahin L, Gelperin K, et al. Utilization of drugs with pregnancy exposure registries during pregnancy. Pharmacoepidemiol Drug Saf. 2018;27(6):604–11. https://doi.org/10.1002/pds.4409.

42. Golder S, Ahmed S, Norman G, Booth A. Attitudes toward the ethics of research using social media: a systematic review. J Med Internet Res. 2017;19(6):e195. https://doi.org/10.2196/jmir.7082.

43. Thorpe PG, Gilboa SM, Hernandez-Diaz S, Lind J, Cragan JD, Briggs G, et al. Medications in the first trimester of pregnancy: most common exposures and critical gaps in understanding fetal risk. Pharmacoepidemiol Drug Saf. 2013;22(9):1013–8. https://doi.org/10.1002/pds.3495.

44. Nezvalova-Henriksen K, Spigset O, Nordeng H. Effects of ibuprofen, diclofenac, naproxen, and piroxicam on the course of pregnancy and pregnancy outcome: a prospective cohort study. BJOG. 2013;120(8):948–59. https://doi.org/10.1111/1471-0528.12192.

45. Li DK, Liu L, Odouli R. Exposure to non-steroidal anti-inflammatory drugs during pregnancy and risk of miscarriage: population based cohort study. BMJ. 2003;327(7411):368. https://doi.org/10.1136/bmj.327.7411.368.

46. Van Marter LJ, Hernandez-Diaz S, Werler MM, Louik C, Mitchell AA. Nonsteroidal antiinflammatory drugs in late pregnancy and persistent pulmonary hypertension of the newborn. Pediatrics. 2013;131(1):79–87. https://doi.org/10.1542/peds.2012-0496.