RESEARCH ARTICLE

Pregnancy and health in the age of the Internet: A content analysis of online "birth club" forums

Anna Wexler[1]*, Anahita Davoudi[2], Davy Weissenbacher[2], Rebekah Choi[1], Karen O'Connor[2], Holly Cummings[3], Graciela Gonzalez-Hernandez[2]

1 Department of Medical Ethics and Health Policy, University of Pennsylvania, Philadelphia, PA, United States of America, **2** Department of Biostatistics, Epidemiology, and Informatics, University of Pennsylvania, Philadelphia, PA, United States of America, **3** Department of Obstetrics and Gynecology, University of Pennsylvania, Philadelphia, PA, United States of America

* awex@pennmedicine.upenn.edu
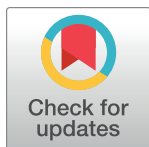
Abstract

Background

Although studies report that more than 90% of pregnant women utilize digital sources to supplement their maternal healthcare, little is known about the kinds of information that women seek from their peers during pregnancy. To date, most research has used self-report measures to elucidate how and why women to turn to digital sources during pregnancy. However, given that these measures may differ from actual utilization of online health information, it is important to analyze the online content pregnant women generate.

Objective

To apply machine learning methods to analyze online pregnancy forums, to better understand how women seek information from a community of online peers during pregnancy.

Methods

Data from seven WhatToExpect.com "birth club" forums (September 2018; January-June 2018) were scraped. Forum posts were collected for a one-year period, which included three trimesters and three months postpartum. Only initial posts from each thread were analyzed (n = 262,238). Automatic natural language processing (NLP) methods captured 50 discussed topics, which were annotated by two independent coders and grouped categorically.

Results

The largest topic categories were maternal health (45%), baby-related topics (29%), and people/relationships (10%). While pain was a popular topic all throughout pregnancy, individual topics that were dominant by trimester included miscarriage (first trimester), labor (third trimester), and baby sleeping routine (postpartum period).

Conclusion

More than just emotional or peer support, pregnant women turn to online forums to discuss their health. Dominant topics, such as labor and miscarriage, suggest unmet informational needs in these domains. With misinformation becoming a growing public health concern, more attention must be directed toward peer-exchange outlets.

Introduction

Pregnant women are increasingly turning to digital information to supplement their maternal health care. One study reported that 97% of women utilize online sources during pregnancy [1]; other research has reported similarly high (>90%) percentages [2–10]. More than just providing emotional support, digital sources influence women's decision-making during pregnancy [1,4,5,11–15]. For example, some women use online information to validate or "get a second opinion" on their physician's advice [1,11,16,17]. Given that online information has a real-world effect on behavior [18] and that women may be seeking Internet health information in lieu of obtaining it from their obstetrician—it is important to understand the kinds of information that women seek during pregnancy.

Prior work in the domain of digital health and pregnancy has focused on elucidating how and why women turn to digital sources during pregnancy [1,3,9,12,13,16,19–23]. However, these studies have a number of limitations. First, most have utilized self-report measures; yet the ways in which women *report* turning to digital health sources may be different than the way they actually *utilize* online health information in practice. Second, most studies have examined information-seeking during pregnancy as a whole, without looking at specific needs that may arise within each trimester. Third, there has been a dearth of research on information-seeking during the postpartum period, which is especially salient given that the American College of Obstetricians and Gynecologists (ACOG) recently identified "the lack of attention to [postpartum] maternal health needs" as an issue of particular concern [24].

An alternate approach to studying how women utilize the Internet during pregnancy is to examine the online content that women generate. Our prior work focused on gathering pregnancy-related information from a generic social media platform such as Twitter, whereby a cohort of pregnant women was identified via their self-reports of pregnancy and their timelines (all publicly available tweets) during pregnancy were analyzed in a case-control study of birth defects [25–27]. Other researchers have identified pregnant women based on their search queries, and determined their time-dependent search queries throughout pregnancy [28].

However, online pregnancy forums—which contain millions of posts generated by pregnant women—have largely gone unexamined. These forums, which are hosted on websites such as WhattoExpect.com and BabyCenter.com, have rapidly increased in popularity in the last decade. The sheer amount of data on these pregnancy forums offers a unique, untapped opportunity to examine how women seek online information from a community of peers during pregnancy. To our knowledge, the only prior general empirical study on these forums was conducted by Gui et al. (2017), who examined 200 online posts in each of three trimester-specific BabyCenter.com forums, utilizing a grounded theory research design [29]. Gui et al. (2017) found that women sought advice, informal and formal knowledge, reassurance, and emotional support; they also identified specific topics of support-seeking across the three trimesters.

At present, the most active discussion boards on pregnancy forums are due-date specific, (also known as "birth clubs"), where expectant mothers can connect with a community of

peers undergoing a similar chronological pregnancy journey.[27] Based on publicly available data from WhatToExpect.com and BabyCenter.com, we estimate that there are currently over one million total posts (initial posts + replies) for each birth club forum (i.e., for each due date month and year, such as April 2019) across these two websites. Though the grounded theory approach utilized by Gui et al. (2017) offered researchers deep immersion in the data, it would be infeasible to manually code the hundreds of thousands of posts that comprise each birth club forum. Automatic and semi-automatic methods of analysis are thus required to utilize the data at scale.

The present study aims to apply automatic methods of language analysis (specifically, Latent Dirichlet allocation (LDA) topic modeling) to a much larger collection of birth club forum postings than what was used by Gui et al. (2017), in order to better understand the kinds of online information that women seek from their peers during pregnancy. We sought to discover whether women discuss health-related topics, and if so, which topics appeared most frequently. In addition, we aimed to differentially characterize the topics that women discuss across the three trimesters and the postpartum period to better understand how women's information needs vary across these time periods.

Methods

Birth club forums

Currently, WhattoExpect.com (owned by j2 Global) and BabyCenter.com (owned by Johnson & Johnson) host the most active online forums for English-speaking women, as determined by number of members and posts. On both WhattoExpect.com and BabyCenter.com, posts are publicly viewable: membership is not required to access the forums, though users that seek to contribute a posting must become members by registering and creating a username. Usernames are not connected to social media accounts; thus the forums provide a level of anonymity. To obtain an approximate measure of the popularity of birth club forums across WhattoExpect.com and BabyCenter.com, we compared the number of total threads appearing in both birth club forums over the past five years (see supplementary material). Though BabyCenter.com was initially more popular, by fall 2018, the number of threads on WhatToExpect.com exceeded those on BabyCenter. Therefore, for the present study we drew our sample from WhatToExpect.com birth club forums. As all data (i.e., posts) for the present study were publicly available, we were advised by the University of Pennsylvania Institutional Review Board (IRB) that the project did not require review.

Birth club dataset

First, we collected data for one WhatToExpect.com birth club forum (September 2018), and then included six other forums (January through June 2018) in our corpus. For each birth club forum, we narrowed our collection to an approximate one-year period that included all three trimesters plus the postpartum period. For our analysis, we considered only initial posts and not data from the entire thread, as initial posts most clearly reflected the goal of the study: to understand the kinds of information that women seek on online forums. Replies to initial posts sometimes digressed and included divergent topics.

In order to distinguish topics throughout the pregnancy from those that arise primarily during a specific trimester, we divided all posts in our corpus into four time periods, corresponding to each of the three trimesters and the postpartum period. Given that we could not know each woman's delivery date, for the purposes of the study we fixed the delivery date to the end of the birth club month, and counted trimesters and the postpartum period in blocks of three months (90 days) back from that date. Thus, for the September 2018 forum, birth was

assumed to have occurred on September 30, and the ensuing 90 days (October 1 through December 29, 2018) was considered as the postpartum period. Working backward in 90-day blocks, posts were divided into the third trimester (July 3 through September 30, 2018), second trimester (April 4 through July 2, 2018), and first trimester (January 4 through April 3, 2018). We chose this normalized division (over using the middle of the month as a due date, for example) as it provided the clearest distinction between the main topics, particularly between the third trimester and the postpartum period. We assigned each post to a unique corresponding trimester based on its date of publication.

This study was deemed exempt from review by the University of Pennsylvania Institutional Review Board due to the fact that it did not meet the regulatory definition of research with "human subjects," as the data was obtained from publicly available sources. In addition, we verified that the study methods were in compliance with the WhatToExpect.com Terms of Use (https://www.whattoexpect.com/terms-of-use/).

Topic modeling and labeling

We utilized automatic natural language processing (NLP) methods to capture the discussed topics. Specifically, we used a form of topic modeling known as Latent Dirichlet Allocation (LDA), which assumes that each document in a large dataset is comprised of sub-topics that are represented by the words they contain. By analyzing the frequency of word appearance in relation to other words, the LDA algorithm can discover a "bag of words" that have a high probability of appearing together. If the model has performed optimally, the bag of words can be identified (or "labelled") by a human coder as relating to a specific topic. Importantly, the topics are not pre-defined, but rather are discovered during the algorithm's training stage.

The performance of topic modeling algorithms is strongly determined by the choices made to pre-process the documents and to optimize the hyper-parameters of the training algorithms. Because some words occur in most documents and do not help to identify distinct topics, we removed the following in the pre-processing stage: (a) all digits; (b) all hyperlinks to external websites; (c) common words such as articles and prepositions (e.g. a, the, of, always, etc.) and (d) generic words related to pregnancy, such as baby, born, days (see supplementary material for complete list of excluded "stop words").

Notably, the number of topics in an LDA model is always finite, and the total number of topics desired is set as an initial parameter. To determine the optimal number of topics for our model, we trained multiple models—with 20, 35, 50, 80 and 150 topics—over all initial posts in the March 2018 and September 2018 forums and compared the resulting sets of words manually. Annotators inspected the word clusters in each set of topics, and chose the 50-topic model given that the word clusters were distinct enough to be assigned a "label" for each topic. With a number of topics lower than 50, the topics failed to capture the full diversity of subjects discussed in the forums. With a number higher than 50, the main topics were not adequately represented as unified topics but rather split into multiple related sub-topics.

The topic modeling system runs through the documents many times during the training phase; topics become more distinct with each additional iteration. We found that 200 iterations were adequate for our model, as additional iterations did not result in noticeable differences in the topics learned. Thus, with these hyper-parameters selected—50 topics and 200 iterations—we computed topics on all initial posts contained in our corpus.

To define topic labels, two annotators independently inspected the 15 words that were the most representative of the resulting 50 topics. Any disagreement in topic labels was resolved through discussion and by reviewing the ten posts most strongly associated with a given topic. A third annotator confirmed the final topic labels through examination of the top 100 posts

associated with each topic, occasionally making minor refinements to the wording of the topic label. Next, all three annotators worked together to group the topics into overall categories and subcategories.

Results

We collected 2,926,324 posts across the seven WhatToExpect.com birth club forums (**Fig 1**). Of those, 262,238 represented initial posts and the remainder were follow-up posts (i.e., all posts following the initial post on a given thread). Initial posts were analyzed (**Fig 1**), for which our model generated 50 distinct word-clusters. Six word-clusters were excluded from our analysis as they were noisy and did not constitute semantically coherent sets. However, annotators successfully labeled 44 word-clusters, and grouped the resulting topics into subcategories and overarching categories (**Fig 2**).

    Fig 3 presents a detailed view of the 44 most dominant topics across all trimesters and their related word clusters. All topics were analyzed by their frequency of appearance using several methods. First, we examined the overall topic frequency across all time periods as a whole (middle bar, **Fig 3**). Next, we analyzed the most frequent topics that appear within each of our four time periods (three trimesters and postpartum). In doing so, we sorted topics from highest-to-lowest in rank order, so that the most frequently-appearing topic in a time period was given a rank of one, the second a rank of two, etc. We analyzed all topics by their rank order across all trimesters (right column, **Fig 3**).

    As shown in **Fig 2**, the largest category of topics was related to maternal health (45%), followed closely by topics related to the baby (29%), and people/relationships (10%). Within the category relating to maternal health, physical symptoms (including nausea, pain, and bleeding) was the most dominant subcategory (**Fig 3**), followed by a subcategory relating to labor (including induction/dilation, labor signs, and anxiety about delivery). Body and mood changes, pregnancy complications, and concerns about miscarriage comprised the remainder of subcategories. When examining the overall count of topics across all time periods, both pain

```
┌─────────────────────────────────┐
│  All posts collected from 7      │
│  different WhatToExpect.com      │
│  2018 birth club months          │
│  (N=2,926,324)                   │
└─────────────────────────────────┘
```

```
┌──────────────────────────┐        ┌─────────────────────────────────┐
│ Total initial posts      │        │ Total comments (n=2,664,086)    │
│ (n=262,238)              │        │  o  Average number of comments  │
│                          │        │     per initial post ± SD: 10 ± 26 │
└──────────────────────────┘        └─────────────────────────────────┘
```

```
┌────────────────────────────────────────────┐
│ Total initial posts analyzed (n=262,238)    │
│ Average per birth club month:               │
│  o  Initial posts: 37,463                   │
│  o  Users (those who made initial posts):   │
│     7,410                                   │
└────────────────────────────────────────────┘
```
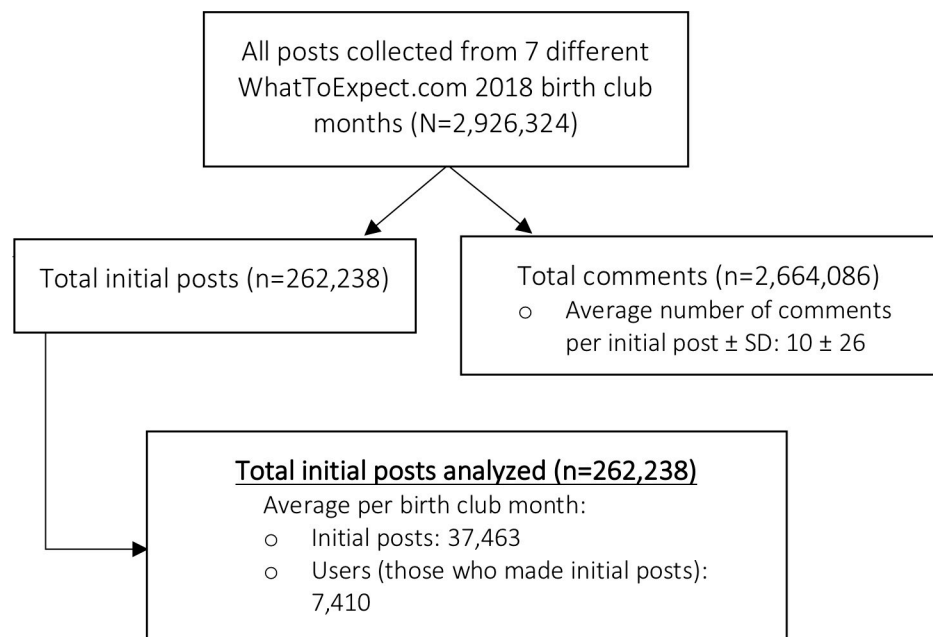
Fig 1. Data collection flowchart.
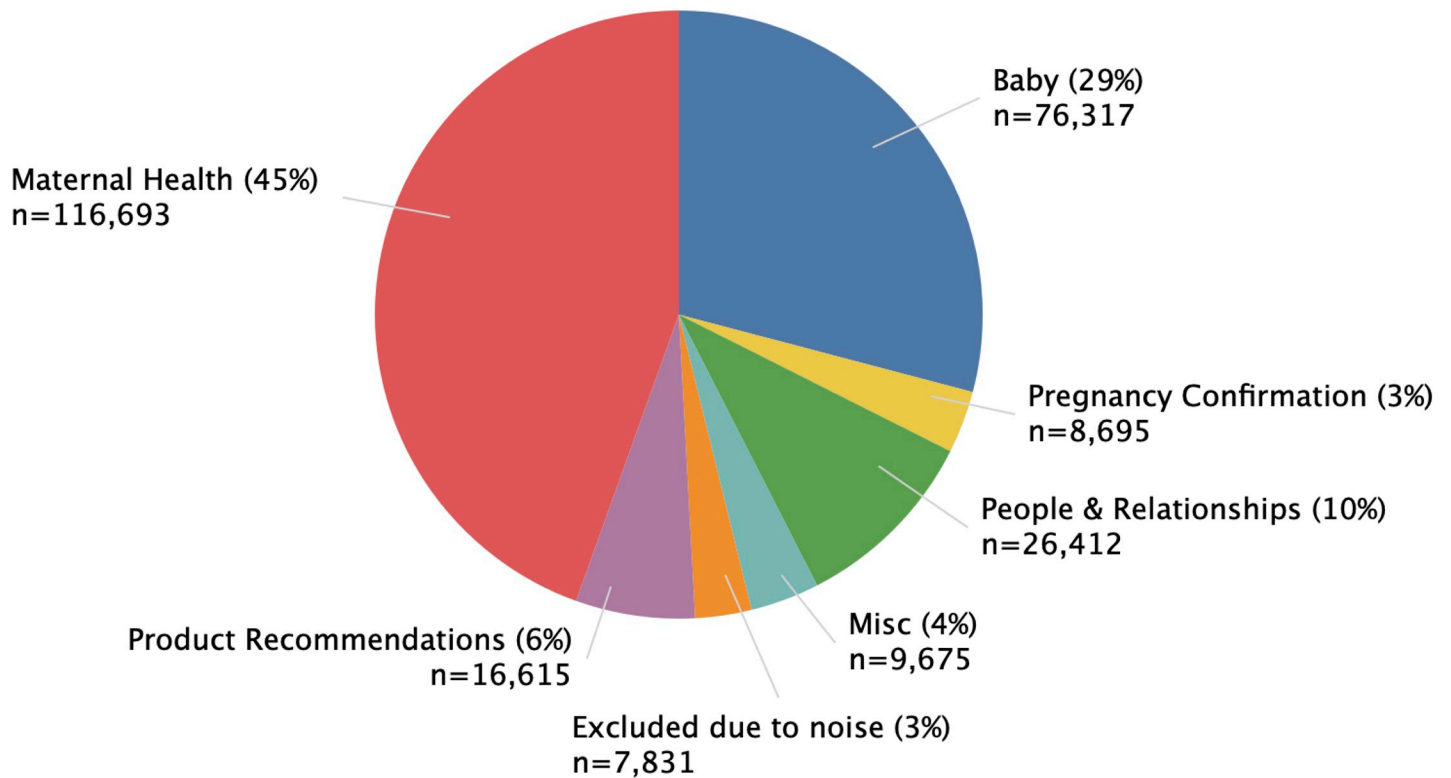
Fig 2. Distribution of posts by overall topic category.

and labor topped the list, as indicated by the dark red squares (middle column, **Fig 3**). An analysis of the topics by rank order across the trimesters (right column, **Fig 3**), shows the dominance of certain topics by trimester: concerns about miscarriage, for example, appear frequently in the first trimester (as indicated by the dark blue square) but are hardly discussed in the ensuing time periods (as reflected by the lighter blue squares). Similarly, topics relating to labor appear most frequently in third trimester, whereas discussions of pain are prevalent across all four time periods.

Within the category relating to babies, caring for the newborn was the most frequent subcategory, which included specific topics relating to baby sleeping and feeding routines, as well as breastfeeding. Overall, "baby sleep routine" was one of the most dominant topics across all time periods (middle bar, **Fig 3**), and topics related to caring for the newborn are ranked highly in the postpartum period (right column, **Fig 3**). Other subcategories relating to baby were preparing for baby (including excitement about baby, ultrasound pictures, and gender reveal) and baby's health in utero (including concerns about ultrasound results and baby movement); these appear most prominently in the first and second trimesters (right column, **Fig 3**).

The category relating to people/relationships only contains two topics: husband and family/friends. However, this category is notable in that these two topics appear frequently overall (middle column, **Fig 3**) as well as frequently within each individual time period (right column, **Fig 3**). Example posts in these topics include questions about when to share a pregnancy announcement, discussions of how a new baby will impact family dynamics, sharing of household labor, and requests for advice on relationship issues.

The category relating to pregnancy confirmation indicates that some women turn to specific birth club forums very early on, to discuss pregnancy test results, to share struggles

Category	Subcategory	Topic Label	Total Posts	Word Cluster	T1	T2	T3	PP
Maternal Health	Physical Symptoms	Pain	13,196	pain, back, sit, hurt, experience, walk, pressure, stomach, stand, pelvic, help, hip				
		Morning sickness	9,751	morning, symptom, nausea, pregnancy, throw, trimester, stomach, sickness, sick, nauseous, vomit, pregnant				
		Bleeding/spotting/cramping	9,357	cramp, bleed, period, spot, blood, discharge, normal, light, experience, red, brown, wipe				
		Misc health symptoms (sore breasts, constipation, hemorrhoids)	5,488	weird, pee, hurt, sore, boob, happen, leak, normal, burn, tmi, heartburn, notice				
		Anxiety & depression; medication questions	4,798	anxiety, pregnancy, help, postpartum, tear, pregnant, infection, issue, doctor, experience, effect, medication				
		Cold & flu symptoms	4,661	sick, cold, cough, flu, fever, shoot, nose, breathe, throat, catch, ear, help				
	Labor	Labor (induction & dilation)	12,062	induce, cm, dilate, labor, push, check, induction, cervix, due, epidural, doctor, bear				
		Labor signs	6,629	contraction, labor, water, break, min, braxton, apart, sign, hick, painful, hospital, strong				
		Anxiety about delivery	5,930	birth, hospital, nervous, scare, delivery, deliver, plan, experience, labor, natural, mom, prepare				
		Mucus plug & baby position questions	1,989	head, show, lose, turn, mucus, plug, bloody, position, breech, flip, flat, mucous				
	Body and mood changes	Diet & food	6,578	eat, drink, food, water, meal, dinner, tea, crave, snack, coffee, lunch, ice				
		Stretch marks & baby bump	5,558	belly, bump, leg, stretch, mark, pop, button, itchy, oil, pregnancy, bloat, itch				
		Weight	4,584	weight, gain, lb, pound, lose, pregnancy, weigh, eat, healthy, pregnant, pre, fat				
		Energy levels & exercise	3,962	pregnancy, body, run, tire, pregnant, walk, struggle, toddler, energy, exhaust, remember, easy				
		Sex drive	1,913	crazy, sex, drive, headache, husband, pregnant, breath, hubby, migraine, pregnancy, nut, insane				
	Pregnancy complications	Pre-eclampsia & doctor's appointments	9,209	doctor, appointment, blood, ob, dr, check, pressure, office, appt, back, nurse, concern				
		Placental disorders	2,132	hear, curious, experience, opinion, placenta, wonder, vs, whether, mum, class, love, mom				
Baby	Preparing for baby	Excited about baby	8,981	girl, boy, excite, love, sweet, meet, beautiful, amaze, healthy, son, team, bless				
		Choosing baby name	8,319	name, middle, girl, boy, love, husband, pick, help, decide, suggestion, choose, opinion				
		First ultrasound picture & early gender prediction	5,343	picture, pic, foot, swell, ramzi, twin, dream, photo, ultrasound, theory, nub, skull				
		Gender reveal	3,171	guess, gender, fun, reveal, csection, surprise, ultrasound, wife, scan, idea, tale, correct				
	Baby's health (in utero)	Concerns about miscarriage	8,896	pregnancy, miscarriage, heartbeat, ultrasound, level, pregnant, hcg, sac, nervous, progesterone, hear, pray				
		Concerns about ultrasound results	8,414	ultrasound, scan, measure, anatomy, doctor, tech, small, growth, experience, result, back, fluid				
		Baby movement	4,625	move, kick, movement, tummy, roll, active, count, back, crawl, normal, lay, doe				
		Baby heart rate	833	heart, rate, beat, glass, cup, wine, program, sippy, babys, proud, bpm, star				
	Caring for newborn	Baby sleep routine	10,573	sleep, night, wake, nap, fall, asleep, back, bed, awake, help, eat, fee				
		Pumping & breastfeeding	9,305	pump, breastfeed, milk, breast, feed, nurse, bottle, supply, nipple, fee, formula, latch				
		Baby feeding routines	5,896	bottle, formula, oz, eat, feed, spit, gas, solid, reflux, fee, burp, switch				
		Baby sleep arrangements	4,505	room, crib, play, sleep, swaddle, bassinet, rock, nursery, transition, arm, bed, pack				
		Skin problems	3,443	change, diaper, face, skin, rash, app, iron, help, wet, acne, red, cream				
		Baby teething & thumb sucking	1,275	hand, mouth, suck, tooth, book, teethe, finger, pull, toy, pacifier, cut, lip				
		Bowel movements	976	poop, dog, video, cat, pooping, potty, green, train, pooped, ride, fresh, normal				
	Misc	Baby eye color	658	eye, open, white, color, blue, black, dark, pink, light, wide, fill, contact				
People & relationships		Husband	13,478	husband, house, kid, work, help, back, care, alone, son, stress, vent, mom				
		Family & friends	12,934	family, friend, mom, child, people, husband, pregnant, mother, sister, parent, law, daughter				
Product recommendations		Buying maternity clothes & baby items	7,325	buy, maternity, bag, size, clothe, stuff, diaper, brand, store, pack, item, newborn				
		Car seats & strollers	6,343	car, seat, job, pay, stroller, insurance, cover, school, money, double, infant, plan				
		Bra recommendations	2,947	wear, clean, wash, fit, bra, dress, bath, clothe, nurse, pad, wrap, free				
Pregnancy confirmation		Pregnancy tests & glucose tests	4,643	test, positive, line, result, negative, blood, pregnancy, glucose, faint, sugar, morning, fail				
		Estimating due date	4,052	due, june, april, september, march, pregnant, january, cycle, bfp, late, august, july				
Misc		Emotional moms & crying babies	4,225	cry, hold, scream, hate, stop, calm, help, smile, love, sad, leap, upset				
		Travel advice	2,758	advice, tip, appreciate, plan, live, travel, trip, help, weekend, fly, husband, vacation				
		Temperature related	1,430	safe, hot, cold, warm, heat, cool, summer, outside, sweat, air, winter, weather				
		Back to work & childcare plans	1,262	work, back, schedule, full, daycare, shift, routine, mom, growth, spurt, hr, overnight				
		Six topics excluded due to noise	7,831					
		TOTAL	262,238					

T1 = First trimester
T2 = Second trimester
T3 = Third trimester
PP = Postpartum period

Rank Per Time Period
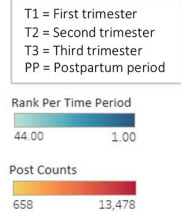44.00   1.00

Post Counts
658   13,478

Fig 3. Topics of discussion on online pregnancy forums. From left to right: Overall topic categories, subcategories, and topic labels. Total posts per topic are visually represented by yellow-red frequency heat map. The word cluster column contains the top 12 words generated by our algorithm, per topic. The teal-blue heat map depicts topics in rank order within each time period.

regarding conception, or to estimate their due date. The category relating to product recommendations shows that most women are looking for advice on clothing (maternity-wear, baby clothes, bras) and baby transport options; the topics here are most dominant in the second trimester, when women are most likely to begin needing maternity clothes and when they start preparing for baby. The final category consists of several miscellaneous topics, regarding emotions, travel advice, temperature, and going back to work. Although the latter topic has only a small number of posts, a significant number of posts in the "husband" category are related to going back to work, as evidenced by the word cluster in that topic and our manual examinations of representative posts; it is therefore likely that our model captured many work-related posts in that category.

Finally, as rank order obscures the numerical differences between the most frequent topics, we also depict the numerical counts of the top 15 ranked posts per time period (**Fig 4**). All time periods except for the second trimester have a topic that clearly dominates. In the first trimester, topics related to concerns about miscarriage, morning sickness, and bleeding/cramping (which is likely also related to miscarriage) appear most frequently. In the third trimester, labor-related topics and pain are the ones that most frequently appear. Notably, not only is "labor (induction/dilation)" the single most frequent topic when looked at in any individual time period, but when considered along with closely related topics in this trimester ("labor signs" and "anxiety about delivery"), discussions of labor total approximately 20,000 posts—an amount that dwarfs all other posts in individual time periods. Finally, two topics—husband and family/friends—appear in the top 15 across all time periods (darker bars, **Fig 4**).

Discussion

The results of this study demonstrate that more than just emotional or peer support, or even product recommendations, pregnant women are turning to online forums to discuss their health. They post about topics such as bleeding, cramping, pain, labor, and morning sickness. They also turn to the forum to discuss worries about their baby's health, such as fears of miscarriage, baby's heart rate, and abnormal ultrasound results. One of the most striking findings from this study is that two topics, miscarriage and labor, are by far the most dominant topics within specific time periods (first and third trimester, respectively).

Our findings are largely distinct from previous work that has directly queried women (via interviews or surveys) about topics of information-seeking during pregnancy, which has found that the most common information sought online during pregnancy is related to fetal development, nutrition in pregnancy, stages of childbirth, and pregnancy complications [1,6,7,9,12,30–35]. Aside from the topic relating to childbirth, which is analogous to our "labor" topic, the remaining topics either did not arise in our findings (i.e, fetal development) or were not dominant in specific trimesters. The topics found in our study, however, are largely similar to those from Gui et al. (2017), who analyzed trimester-specific forums on BabyCenter.com, and overlap to some extent with a study examining pregnant women's search queries [28].

Thus, it seems that the few studies examining the actual content of either forum conversations or search queries have largely converged on similar findings, whereas studies that have asked women about their online information-seeking behavior have resulted in a second set of findings. There are several possible explanations for this divergence. First, there may be
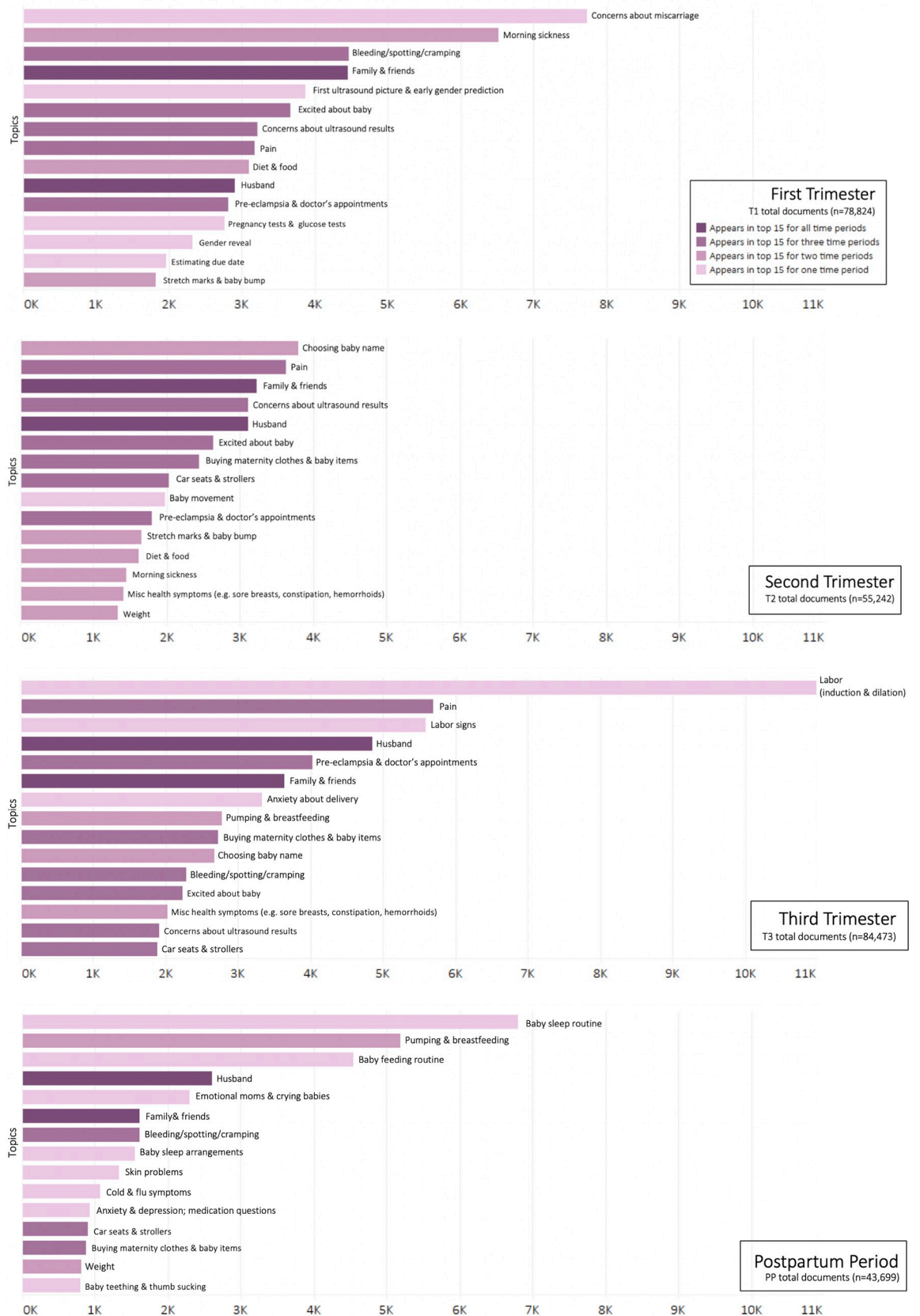
First Trimester
T1 total documents (n=78,824)
■ Appears in top 15 for all time periods
■ Appears in top 15 for three time periods
■ Appears in top 15 for two time periods
■ Appears in top 15 for one time period

Topics (First Trimester): Concerns about miscarriage; Morning sickness; Bleeding/spotting/cramping; Family & friends; First ultrasound picture & early gender prediction; Excited about baby; Concerns about ultrasound results; Pain; Diet & food; Husband; Pre-eclampsia & doctor's appointments; Pregnancy tests & glucose tests; Gender reveal; Estimating due date; Stretch marks & baby bump

Second Trimester
T2 total documents (n=55,242)

Topics (Second Trimester): Choosing baby name; Pain; Family & friends; Concerns about ultrasound results; Husband; Excited about baby; Buying maternity clothes & baby items; Car seats & strollers; Baby movement; Pre-eclampsia & doctor's appointments; Stretch marks & baby bump; Diet & food; Morning sickness; Misc health symptoms (e.g. sore breasts, constipation, hemorrhoids); Weight

Third Trimester
T3 total documents (n=84,473)

Topics (Third Trimester): Labor (induction & dilation); Pain; Labor signs; Husband; Pre-eclampsia & doctor's appointments; Family & friends; Anxiety about delivery; Pumping & breastfeeding; Buying maternity clothes & baby items; Choosing baby name; Bleeding/spotting/cramping; Excited about baby; Misc health symptoms (e.g. sore breasts, constipation, hemorrhoids); Concerns about ultrasound results; Car seats & strollers

Postpartum Period
PP total documents (n=43,699)

Topics (Postpartum Period): Baby sleep routine; Pumping & breastfeeding; Baby feeding routine; Husband; Emotional moms & crying babies; Family & friends; Bleeding/spotting/cramping; Baby sleep arrangements; Skin problems; Cold & flu symptoms; Anxiety & depression; medication questions; Car seats & strollers; Buying maternity clothes & baby items; Weight; Baby teething & thumb sucking

genuine differences between the kinds of information that women glean from apps or static websites as compared to what they post on online forums. For example, if an information need, such as stages of fetal development, is satisfied by an app, women may be less inclined to post questions about it on an online forum. Second, the interview and survey studies varied in query methodology. Women may have responded differently if a question encompassed use of digital sources as a whole, as compared to a question about Internet searches. Third, there may have been a recall bias in interview and survey studies, resulting in participants not accurately recalling their most commonly searched topics. Fourth, there could be a selection bias in that women who post on online forums are interested in different topics than a random sample of pregnant women.

Notably, pregnant women are likely encountering forum discussions even if they have not joined an online community. As noted above, the vast majority of women are turning to online information during pregnancy, and the way they do so is primarily through the use of search engines [1,3,5,6,11,19,36]. While generic search terms such as "pregnancy" or "miscarriage" may yield results from static websites, natural language queries such as "is shooting pain during the second trimester normal" may generate links to forum posts. Indeed, one study that analyzed search engine results for pregnancy-related queries found that natural language search terms (i.e., "my baby is moving less") mostly resulted in forum posts, whereas general terms such as "fetal movement" mostly generated results for static webpages and articles [37]. In other words, women who use natural language terms in their search queries will likely stumble across forum posts at some point—or possibly even quite frequently—during their pregnancy.

This is especially concerning, given that as many as 70–75% of pregnant women do not speak to their health care providers about information retrieved from the Internet [23,30,34]. Furthermore, one recent study analyzing the quality of online information in Internet discussion forums in pregnancy found that 24.3% of responses lacked credible evidence and 5.5% were potentially harmful [38]. Yet women typically perceive online pregnancy information to be reliable and of high quality [1,3,5,9,30,34], and use it to inform their health care decision-making [4,5,12,14]. Thus, while information retrieved from online forums may be influencing women's health choices, this behavior may not always be transparent to the health care provider.

To the extent that the most dominant topics in **Fig 3** can be interpreted as unmet informational needs, our study has a number of practical implications. First, pregnancy literature typically provides women with information about common symptoms and body changes, such as weight gain, nausea, stretch marks, abdominal and back pain, and energy levels. There is little information, other than "call your doctor," regarding pregnancy complications, such as miscarriage, pre-eclampsia, or abnormal ultrasound results [17]. On the one hand, given that the majority of women will not experience these complications, this strategy may avoid causing undue concern. On the other hand, in the age of the Internet, the lack of readily available information about "atypical" pregnancies ensures that those experiencing complications will likely go online to search for information. Women may do this either in lieu of contacting their health providers, to prepare for appointments, or even to "double-check" information provided to them by their physician [1,11,16]. Health care providers should recognize that they may no longer be the primary source of health information during pregnancy, and professional societies may want to consider providing reliable information online regarding pregnancy complications.

Second, the total number of posts related to labor points to a critical information gap in this area. Currently, aside from optional childbirth classes, there is little to prepare women for labor, both from an emotional or informational perspective. Labor is perhaps the area that generates the most uncertainty and anxiety during pregnancy, especially for first-time mothers: specific posts in our sample included questions about how to discern when one is in labor and/or when one's water has broken; when to go the hospital; how to induce labor; and how to cope when a birth is proceeding differently than expected. Given the significant informational and support needs related to labor, it is not surprising that studies have consistently found benefits of companionship during the labor process (e.g., such as that provided by a doula) [39–41]. The results of our study, which indicate that there is still a significant unmet informational and emotional support need, support ACOG's recommendation of incorporating support personnel such as doulas during labor, as well as other suggestions to have labor support routinely provided by healthcare systems [41–43].

Third, the number of topics in the postpartum period related to newborn care—particularly with regard to feeding and sleeping routines—indicates another crucial information gap. Here, too, there is often little information about newborn care provided to women in the lead-up to birth; after delivery, women must shift their questions to a new care provider, the pediatrician. Indeed, in the immediate postpartum period—a time when women are experiencing significant physical and psychological challenges—the provider who has supported them throughout pregnancy is no longer responsible for care, and will not typically see them for six weeks. From the women's perspective, pregnancy, labor and newborn care are a continuous process; this continuity, however, is not mirrored by health care systems [44]. The results from our study suggest that additional information and emotional support should be provided to women during the postpartum period.

This study has a number of limitations. The topics in **Fig 2** do not indicate *all* the topics discussed during pregnancy but rather indicate word clusters that our model determined had the strongest correlations between words. Posts that were very general in nature, or those that did not include unique words, may not have been detected by our model. In addition, some word clusters may not have arisen due to our list of excluded words (e.g., the exclusion of "born" and "section" likely precluded word clusters related to birth announcements and caesarian sections). Furthermore, given that our model is based on associations between words, some of our topics contain multiple sub-topics with related words. For example, the topic related to preeclampsia and blood pressure also contained general posts related to doctor and doctor's appointments, and the topic related to bowel movements was mostly related to baby but also contained posts regarding exposure to cat feces during pregnancy. Finally, as noted in the methods section, we could not capture the exact date of birth for each poster, thus we relied on estimates of trimester timeframes.

Because WhatToExpect.com displays limited information about each user, we could not determine users' genders, locations, or ages. Nor could we determine the relative percentages of nulliparous and primiparous women. All posts we reviewed appeared to be from heterosexual women, which is in-line with the findings of Gui et. al (2017)'s analysis of the BabyCenter forum. It was also our impression that the majority of users are from the United States, though some users appear to be living in Canada, the United Kingdom, and Australia, as evidenced by the usage of certain words (e.g., "pram" and "nappy") and mentions of different health care systems.

To our knowledge, this study is the first to provide a large-scale, big data analysis of online pregnancy forums. While several previous studies have manually examined forum posts relating to specific topics—such as vaginal birth after caesarean section [45], multiple sclerosis [46], and vaginal breech birth [47]—the present study examined 262,238 posts from seven different

birth club months, yielding a general topical portrait of all posts. By conducting frequency analyses both overall and within four different timeframes, we were able to highlight dominant topics overall as well as those that were trimester-specific.

In some sense, the fact that individuals are utilizing online forums to seek out health information is nothing new [18]. In the realm of pregnancy, however, the data is clear: the vast majority of women are utilizing digital health information during pregnancy and are using that information to inform their decision-making, all while not necessarily mentioning this process (or retrieved information) to their healthcare provider. It is crucial that more scholarly attention be directed to peer-exchange outlets such as online pregnancy forums, especially as "misinformation" is becoming increasingly recognized as a matter of public health importance [48–50]. Professional societies may want to consider providing guidelines to health care providers regarding how to navigate their patients' use of online information. This study, which outlines the topics of discussion on online pregnancy forums, should provide the first step in informing that pathway forward.

Supporting information

S1 Data.
(PDF)

Author Contributions

Conceptualization: Anna Wexler, Anahita Davoudi, Davy Weissenbacher, Graciela Gonzalez-Hernandez.

Data curation: Anahita Davoudi, Davy Weissenbacher, Graciela Gonzalez-Hernandez.

Formal analysis: Anna Wexler, Anahita Davoudi, Davy Weissenbacher, Karen O'Connor, Graciela Gonzalez-Hernandez.

Funding acquisition: Anna Wexler, Graciela Gonzalez-Hernandez.

Investigation: Anna Wexler, Graciela Gonzalez-Hernandez.

Methodology: Anna Wexler, Anahita Davoudi, Davy Weissenbacher, Graciela Gonzalez-Hernandez.

Project administration: Anna Wexler, Graciela Gonzalez-Hernandez.

Resources: Anna Wexler, Graciela Gonzalez-Hernandez.

Software: Graciela Gonzalez-Hernandez.

Supervision: Anna Wexler, Davy Weissenbacher, Graciela Gonzalez-Hernandez.

Validation: Holly Cummings, Graciela Gonzalez-Hernandez.

Visualization: Anna Wexler, Rebekah Choi, Graciela Gonzalez-Hernandez.

Writing – original draft: Anna Wexler, Davy Weissenbacher.

Writing – review & editing: Anna Wexler, Rebekah Choi, Holly Cummings, Graciela Gonzalez-Hernandez.

References

1. Lagan BM, Sinclair M, Kernohan GW. Internet Use in Pregnancy Informs Women's Decision Making: A Web-Based Survey. Birth. 2010; 37: 106–115. https://doi.org/10.1111/j.1523-536X.2010.00390.x PMID: 20557533

2.   Bjelke M, Martinsson A-K, Lendahls L, Oscarsson M. Using the Internet as a source of information dur-
     ing pregnancy—A descriptive cross-sectional study in Sweden. Midwifery. 2016; 40: 187–191. https://
     doi.org/10.1016/j.midw.2016.06.020 PMID: 27450590

3.   Bert F, Gualano M, Brusaferro S, Vito E, de Waure C, Torre G, et al. Pregnancy e-health: a multicenter
     Italian cross-sectional study on internet use and decision-making among pregnant women. J Epidemiol
     Commun H. 2013; 67: 1013. https://doi.org/10.1136/jech-2013-202584 PMID: 24072743

4.   Slomian J, Bruyère O, Reginster JY, Emonts P. The internet as a source of information used by women
     after childbirth to meet their need for information: A web-based survey. Midwifery. 2017; 48: 46–52.
     https://doi.org/10.1016/j.midw.2017.03.005 PMID: 28324809

5.   Huberty J, Dinkel D, Beets MW, Coleman J. Describing the Use of the Internet for Health, Physical
     Activity, and Nutrition Information in Pregnant Women. Matern Child Healt J. 2013; 17: 1363–1372.
     https://doi.org/10.1007/s10995-012-1160-2 PMID: 23090284

6.   Lima-Pereira P, Bermúdez-Tamayo C, Jasienska G. Use of the Internet as a source of health informa-
     tion amongst participants of antenatal classes. J Clin Nurs. 2012; 21: 322–330. https://doi.org/10.1111/
     j.1365-2702.2011.03910.x PMID: 22093043

7.   Sayakhot P, Carolan-Olah M. Internet use by pregnant women seeking pregnancy-related information:
     a systematic review. Bmc Pregnancy Childb. 2016; 16: 65. https://doi.org/10.1186/s12884-016-0856-5
     PMID: 27021727

8.   Declercq ER, Sakala C, Corry MP, Applebaum S, Herrlich A. Listening to Mothers III Pregnancy and
     Birth: Report of the Third National U.S. Survey of Women's Childbearing Experiences. 2013;

9.   Jacobs E, van Steijn ME, van Pampus MG. Internet usage of women attempting pregnancy and preg-
     nant women in the Netherlands. Sex Reprod Healthc. 2019; 21: 9–14. https://doi.org/10.1016/j.srhc.
     2019.04.005 PMID: 31395240

10. Lewallen L, Côté-Arsenault DY. Implications for Nurses and Researchers of Internet Use by Childbear-
     ing Women. Nurs Women's Heal. 2014; 18: 392–400. https://doi.org/10.1111/1751-486x.12147 PMID:
     25316539

11. Lagan BM, Sinclair M, Kernohan GW. What Is the Impact of the Internet on Decision-Making in Preg-
     nancy? A Global Study. Birth. 2011; 38: 336–345. https://doi.org/10.1111/j.1523-536X.2011.00488.x
     PMID: 22112334

12. Narasimhulu D, Karakash S, Weedon J, Minkoff H. Patterns of Internet Use by Pregnant Women, and
     Reliability of Pregnancy-Related Searches. Matern Child Healt J. 2016; 20: 2502–2509. https://doi.org/
     10.1007/s10995-016-2075-0 PMID: 27456311

13. Song F, West J, Lundy L, Dahmen N. Women, Pregnancy, and Health Information Online. Gender Soc.
     2012; 26: 773–798. https://doi.org/10.1177/0891243212446336

14. Ford AJ, Alwan NA. Use of social networking sites and women's decision to receive vaccinations during
     pregnancy: A cross-sectional study in the UK. Vaccine. 2018; 36: 5294–5303. https://doi.org/10.1016/j.
     vaccine.2018.07.022 PMID: 30055969

15. Nikolova G, Lynch C. Do mothers use the internet for pregnancy-related information and does it affect
     their decisions during the pregnancy? A literature review. MIDIRS Midwifery Digest. 2015; 25.

16. Prescott J, Mackie L. Correction: "You Sort of Go Down a Rabbit Hole . . .You're Just Going to Keep on
     Searching": A Qualitative Study of Searching Online for Pregnancy-Related Information During Preg-
     nancy. J Med Internet Res. 2017; 19: e223. https://doi.org/10.2196/jmir.8227 PMID: 30578202

17. Lowe P, Powell J, Griffiths F, Thorogood M, Locock L. "Making it All Normal": The Role of the Internet in
     Problematic Pregnancy. Qual Health Res. 2009; 19: 1476–1484. https://doi.org/10.1177/
     1049732309348368 PMID: 19805809

18. Fox S, Duggan M. Health Online 2013. Pew Research Center. 2013;

19. Lupton D. The use and value of digital media for information about pregnancy and early motherhood: a
     focus group study. Bmc Pregnancy Childb. 2016; 16: 171. https://doi.org/10.1186/s12884-016-0971-3
     PMID: 27435182

20. Cohen JH, Raymond JM. How the Internet is Giving Birth (To) a New Social Order. Information Com-
     mun Soc. 2011; 14: 937–957. https://doi.org/10.1080/1369118x.2011.582132

21. Romano AM, Romano AM. A Changing Landscape: Implications of Pregnant Women's Internet Use for
     Childbirth Educators. Romano Amy M. 2007; 16: 18–24(7). https://doi.org/10.1624/105812407x244903
     PMID: 18769519

22. Javanmardi M, Noroozi M, Mostafavi F, Ashrafi-Rizi H. Internet Usage among Pregnant Women for
     Seeking Health Information: A Review Article. Iranian J Nurs Midwifery Res. 2018; 23: 79–86. https://
     doi.org/10.4103/ijnmr.ijnmr_82_17 PMID: 29628953

23. Fredriksen E, Harris J, Moland K. Web-based Discussion Forums on Pregnancy Complaints and Maternal Health Literacy in Norway: A Qualitative Study. J Med Internet Res. 2016; 18: e113. https://doi.org/10.2196/jmir.5270 PMID: 27230094

24. Committee ACOG. Optimizing postpartum care. ACOG Committee Opinion 736. Obstetrics and Gynecology. 2018; e140–50. https://doi.org/10.1097/AOG.0000000000002633

25. Klein AZ, Sarker A, Weissenbacher D, Gonzalez-Hernandez G. Towards scaling Twitter for digital epidemiology of birth defects. Npj Digital Medicine. 2019; 2: 96. https://doi.org/10.1038/s41746-019-0170-5 PMID: 31583284

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; 87: 68–78. https://doi.org/10.1016/j.jbi.2018.10.001 PMID: 30292855

27. Golder S, Chiuve S, Weissenbacher D, Klein A, O'Connor K, Bland M, et al. Pharmacoepidemiologic Evaluation of Birth Defects from Health-Related Postings in Social Media During Pregnancy. Drug Safety. 2019; 42: 389–400. https://doi.org/10.1007/s40264-018-0731-6 PMID: 30284214

28. Fourney A, White RW, Horvitz E. Exploring Time-Dependent Concerns about Pregnancy and Childbirth from Search Logs. 2015; 737–746. https://doi.org/10.1145/2702123.2702427

29. Gui X, Chen Y, Kou Y, Pine K, Chen Y. Investigating Support Seeking from Peers for Pregnancy in Online Health Communities. Proc Acm Human-computer Interact. 2017; 1: 50. https://doi.org/10.1145/3134685

30. Larsson M. A descriptive study of the use of the Internet by women seeking pregnancy-related information. Midwifery. 2009; 25: 14–20. https://doi.org/10.1016/j.midw.2007.01.010 PMID: 17408822

31. Kavlak O, Atan Ş, Güleç D, Öztürk R, Atay N. Pregnant women's use of the internet in relation to their pregnancy in Izmir, Turkey. Informatics Heal Soc Care. 2012; 37: 253–263. https://doi.org/10.3109/17538157.2012.710686 PMID: 22958087

32. Almoajel A, Almarqabi N. Online Health-information Seeking Behavior among Pregnant Women in Prenatal Clinics at King Saud Medical City, Riyadh. J Womens Heal Issues Care. 2016; 2016. https://doi.org/10.4172/2325-9795.1000228

33. Demirci J, Cohen SM, Parker M, Holmes A, Bogen DL. Access, Use, and Preferences for Technology-Based Perinatal and Breastfeeding Support Among Childbearing Women. J Périnat Educ. 2016; 25: 29–36(8). https://doi.org/10.1891/1058-1243.25.1.29 PMID: 26848248

34. Gao L, Larsson M, Luo S. Internet use by Chinese women seeking pregnancy-related information. Midwifery. 2013; 29: 730–735. https://doi.org/10.1016/j.midw.2012.07.003 PMID: 22958935

35. Scaioli G, Bert F, Galis V, Brusaferro S, Vito DE, Torre LG, et al. Pregnancy and internet: sociodemographic and geographic differences in e-health practice. Results from an Italian multicenter study. Public Health. 2015; 129: 1258–1266. https://doi.org/10.1016/j.puhe.2015.06.012 PMID: 26210071

36. Rodger D, Skuse A, Wilmore M, Humphreys S, Dalton J, Flabouris M, et al. Pregnant women's use of information and communications technologies to access pregnancy-related health information in South Australia. Aust J Prim Health. 2013; 19: 308–312. https://doi.org/10.1071/PY13029 PMID: 24004661

37. Farrant K, Heazell A. Online information for women and their families regarding reduced fetal movements is of variable quality, readability and accountability. Midwifery. 2016; 34: 72–78. https://doi.org/10.1016/j.midw.2015.12.013 PMID: 26867705

38. Ellis L, Roberts L. Exploring the use and quality of Internet discussion forums in pregnancy: A qualitative analysis. Birth. 2019; https://doi.org/10.1111/birt.12459 PMID: 31583769

39. Bohren MA, Hofmeyr JG, Sakala C, Fukuzawa RK, Cuthbert A. Continuous support for women during childbirth. Cochrane Db Syst Rev. 2017; 7: CD003766. https://doi.org/10.1002/14651858.cd003766.pub6 PMID: 28681500

40. Bohren MA, Berger BO, Munthe-Kaas H, Tunçalp Ö. Perceptions and experiences of labour companionship: a qualitative evidence synthesis. Cochrane Db Syst Rev. 2019; 3: CD012449. https://doi.org/10.1002/14651858.cd012449.pub2 PMID: 30883666

41. Lunda P, Minnie C, Benadé P. Women's experiences of continuous support during childbirth: a meta-synthesis. Bmc Pregnancy Childb. 2018; 18: 167. https://doi.org/10.1186/s12884-018-1755-8 PMID: 29764406

42. ACOG Committee. ACOG Committee Opinion No. 766 Summary. Obstetrics Gynecol. 2019; 133: 406–408. https://doi.org/10.1097/aog.0000000000003081 PMID: 30681540

43. Dahlen HG. It is time to consider labour companionship as a human rights issue. Évid Based Nurs. 2019; ebnurs-2019-103127. https://doi.org/10.1136/ebnurs-2019-103127 PMID: 31515238

44. Lundgren I. Swedish women's experiences of doula support during childbirth. Midwifery. 2010; 26: 173–180. https://doi.org/10.1016/j.midw.2008.05.002 PMID: 18639959

45. Konheim-Kalkstein YL, Whyte R, Miron-Shatz T, Stellmack MA. What are VBAC Women Seeking and Sharing? A Content Analysis of Online Discussion Boards. Birth. 2015; 42: 277–282. https://doi.org/10.1111/birt.12167 PMID: 26032668

46. Rezaallah B, Lewis D, Pierce C, Zeilhofer H-F, Berg B-I. Social Media Surveillance of Multiple Sclerosis Medications Used During Pregnancy and Breastfeeding: Content Analysis. J Med Internet Res. 2019; 21: e13003. https://doi.org/10.2196/13003 PMID: 31392963

47. Petrovska K, Sheehan A, Homer C. The fact and the fiction: A prospective study of internet forum discussions on vaginal breech birth. Women Birth. 2017; 30: e96–e102. https://doi.org/10.1016/j.wombi.2016.09.012 PMID: 28277311

48. Chou W-Y, Oh A, Klein WM. Addressing Health-Related Misinformation on Social Media. Jama. 2018; 320: 2417. https://doi.org/10.1001/jama.2018.16865 PMID: 30428002

49. Wang Y, McKee M, Torbica A, Stuckler D. Systematic literature review on the spread of health-related misinformation on social media. Soc Sci Med. 2019; 240: 112552. https://doi.org/10.1016/j.socscimed.2019.112552 PMID: 31561111

50. Armstrong PW, Naylor DC. Counteracting Health Misinformation. Jama. 2019; 321: 1863–1864. https://doi.org/10.1001/jama.2019.5168 PMID: 31009036