

From "Mom Groups" to Medical Triage: A Multi-Methodological Analysis of Pre-eclampsia Discourse on Reddit (2012–2025)

Ananya Singla, Pratham Arora
Computational Social Science
Plaksha University

December 10, 2025

Abstract

Background: Pre-eclampsia is a hypertensive disorder of pregnancy requiring vigilant monitoring. As digital health communities mature, patients increasingly turn to social media for peer support and "peer-led" triage. **Methods:** We engineered a targeted Python collection pipeline (using the PRAW library) targeting 19 subreddits (17 yielded relevant results). Using a dataset of **2,022 posts** and **19,887 comments** collected from June 29, 2013 (when `r/preeclampsia` was founded), we applied VADER sentiment analysis, Latent Dirichlet Allocation (LDA) topic modeling, and TF-IDF keyword extraction to identify linguistic shifts over an 11-year period. **Results:** The analysis reveals a "Hockey Stick" volume expansion beginning in 2020 with COVID-19 as the catalyst, accelerating dramatically through 2023–2025. We observed a "Great Migration" from generalist parenting forums to the specialized `r/preeclampsia` subreddit (54.7% of total volume). Crucially, we identified a "Buffer Effect" where the community transforms negative distress signals (48.8% of posts) into positive support (66.2% of comments). **Conclusion:** The online pre-eclampsia community has evolved from a social outlet into a high-volume, peer-triage ecosystem. However, mass adoption has come with a "sentiment cost," as the density of positive discourse dropped from 57.8% in 2023 to 46.8% in 2025.

1 Introduction

1.1 Clinical Context and Public Health Significance

Pre-eclampsia represents one of the most clinically challenging hypertensive disorders of pregnancy, affecting 2–8% of pregnancies globally and contributing to approximately 76,000 maternal deaths annually worldwide. This condition is characterized by new-onset hypertension (blood pressure $\geq 140/90$ mmHg) and proteinuria (protein in urine) occurring after 20 weeks of gestation. The syndrome can rapidly progress to life-threatening

complications including HELLP syndrome (Hemolysis, Elevated Liver enzymes, Low Platelet count), eclampsia (seizures), placental abruption, and organ failure.

What makes pre-eclampsia particularly insidious from a patient experience perspective is its symptomatology. Unlike conditions with clear diagnostic markers, pre-eclampsia often presents with subjective, easily-dismissed symptoms: headaches, visual disturbances, epigastric pain, and edema, symptoms that overlap significantly with normal pregnancy discomforts. This diagnostic ambiguity creates a critical window where patients must decide whether their symptoms warrant immediate medical attention or are simply "normal pregnancy complaints." It is precisely within this decision-making gap that online health communities have emerged as a crucial informal triage mechanism.

1.2 The Digital Health Revolution and Reddit as a Research Platform

The past decade has witnessed an unprecedented shift in health information-seeking behavior. Traditional gatekeepers of medical knowledge, such as physicians, textbooks, and institutional health websites, now compete with peer-to-peer knowledge exchange platforms. Reddit, with its 52 million daily active users and thousands of health-focused communities (subreddits), has emerged as a particularly important venue for this phenomenon.

Reddit's unique architecture offers several advantages for studying patient discourse:

1. **Pseudonymity:** Users can discuss sensitive health conditions with reduced stigma compared to identifiable platforms like Facebook.
2. **Archival Permanence:** Unlike ephemeral platforms (Instagram Stories, Snapchat), Reddit posts remain searchable indefinitely, enabling longitudinal analysis.
3. **Community Moderation:** Subreddit-specific rules and moderator oversight create relatively structured discourse compared to unmoderated forums.
4. **Temporal Metadata:** Precise timestamps enable temporal analysis of discourse evolution and response to external events (e.g., COVID-19 pandemic).

1.3 Gap in Current Literature

While the clinical pathophysiology of pre-eclampsia is extensively documented, the digital epidemiology of the condition, specifically how patients discuss, self-diagnose, emotionally process, and seek peer support for it online, remains critically understudied. Existing literature on online health communities has largely focused on chronic conditions (diabetes, cancer support groups) or mental health forums (depression, anxiety subreddits). Pregnancy-related conditions, particularly acute and time-sensitive ones like pre-eclampsia, have received minimal scholarly attention despite their prevalence and severity.

Furthermore, most studies of online health discourse employ simple keyword search or manual content analysis of small samples. Few have implemented domain-aware, weighted

sampling strategies that account for the signal-to-noise ratio differences between specialized and generalist health forums.

1.4 Research Questions and Hypotheses

This study argues that the online pre-eclampsia ecosystem has fundamentally transformed from casual social sharing to a structured, high-volume medical peer-triage system. We address three critical research questions:

- **RQ1 (Temporal Evolution):** How has the volume and nature of pre-eclampsia discourse evolved from the pre-pandemic era (2012–2019) to the post-pandemic digital expansion (2023–2025)?

Hypothesis: We expected to observe an immediate volume spike during the COVID-19 pandemic (2020–2021) due to general health anxiety and reduced in-person healthcare access. However, our data reveals a more complex pattern: a delayed “hockey stick” explosion beginning in 2023, suggesting that digital health adoption for pregnancy conditions follows a distinct trajectory from general pandemic-driven telemedicine adoption.

- **RQ2 (Spatial Migration):** How has the locus of discussion migrated from generalist parenting subreddits (`r/BabyBumps`, `r/pregnant`) to dedicated specialized forums (`r/preeclampsia`) over the last decade?

Hypothesis: We anticipated gradual centralization as the condition-specific subreddit matured. Our findings confirm extreme centralization, with `r/preeclampsia` now accounting for 54.7% of all discourse, a dramatic shift from the fragmented landscape of the pre-pandemic era.

- **RQ3 (Community Function):** How effectively does the community function as an emotional buffer, as measured by the sentiment gap between initial distress signals (posts) and community responses (comments)?

Hypothesis: We predicted that posts would exhibit higher negative sentiment (reflecting acute distress) while comments would demonstrate supportive, positive sentiment. Our data reveals a striking “sentiment inversion”: 48.8% of posts are negative, while 66.2% of comments are positive, a 17.4 percentage point gap demonstrating robust community buffering.

1.5 Contribution to Knowledge

This research makes four primary contributions:

1. **Methodological Innovation:** We introduce a weighted, LLM-validated keyword filtering system that dramatically improves signal-to-noise ratio in health discourse mining.
2. **Temporal Insight:** We challenge the prevailing assumption that COVID-19 was the primary driver of digital health adoption, instead identifying 2023 as the inflection point for pregnancy-related online communities.

3. **Structural Analysis:** We document the "Great Migration" from generalist to specialist forums, quantifying the consolidation of expertise into dedicated communities.
4. **Affective Dynamics:** We empirically demonstrate the "Buffer Effect," the systematic transformation of individual distress into collective support, through large-scale sentiment analysis.

2 Methodology

2.1 Data Collection Architecture

2.1.1 Subreddit Selection and Weighting Strategy

Rather than employing uniform sampling across all pregnancy-related subreddits, we developed a domain-aware weighting system that accounts for signal quality differences across forum types. We targeted 19 subreddits, of which 17 yielded relevant data and were included in the final corpus. These were categorized into three tiers:

Tier 1 - Dedicated Condition Forums (Weight = 1.0):

- r/preeclampsia (10,000 expected posts)
- r/NICUParents (5,000 posts) - captures severe outcomes

Tier 2 - High-Relevance Pregnancy Forums (Weight = 0.6–0.8):

- r/BabyBumps, r/pregnant, r/PregnancyAfterLoss
- These capture the "diagnostic moment" when general pregnancy discussion shifts to condition-specific concern

Tier 3 - Broad Medical/Parenting Forums (Weight = 0.3–0.5):

- r/AskDocs, r/Mommit, r/beyondthebump
- Included to capture incidental mentions and general awareness

Mathematical Implementation: The weight w_s for subreddit s determines the post collection quota:

$$N_s = \text{BaseQuota} \times w_s \times \left(1 + \frac{f_s}{f_{max}}\right) \quad (1)$$

where f_s is the observed keyword frequency in subreddit s , and f_{max} is the maximum frequency across all subreddits. This adaptive formula ensures we don't oversample low-signal forums while still capturing the full diversity of discourse.

Rationale: In general parenting forums, the term "preeclampsia" often appears in low-value contexts: healthy users expressing generic pregnancy anxiety ("I'm scared I'll get

preeclampsia”), or off-topic mentions (“My sister had preeclampsia”). In contrast, every post in `r/preeclampsia` represents high-signal discourse from users with direct condition experience. Our weighting prevents dataset dilution while preserving the crucial “diagnostic moment” when users first transition from general pregnancy concerns to condition-specific investigation.

2.1.2 LLM-Validated Keyword Taxonomy

We developed a hierarchical keyword taxonomy validated through iterative consultation with Claude, Gemini, and GPT-4, creating consensus-based medical terminology lists:

Category 1: Core Medical Terms (High Specificity)

- Formal diagnoses: “preeclampsia”, “pre-eclampsia”, “eclampsia”, “HELLP”, “toxemia”
- Clinical markers: “proteinuria”, “gestational hypertension”, “pregnancy-induced hypertension”
- Diagnostic thresholds: “140/90”, “protein in urine”, “elevated liver enzymes”

Category 2: Clinical Symptoms (Medium Specificity)

- Neurological: “vision changes”, “blurred vision”, “seeing spots”, “severe headache”
- Cardiovascular: “high blood pressure”, “hypertension”, “bp readings”
- Other: “epigastric pain”, “right upper quadrant pain”, “facial swelling”

Category 3: Contextual Monitoring Terms (Variable Specificity)

- Self-monitoring: “blood pressure monitor”, “bp cuff”, “readings”, “checking my pressure”
- Medical interaction: “ob said”, “doctor ordered”, “24-hour urine”, “induced early”

Filtering Logic: For Tier 1 subreddits (dedicated forums), we applied no keyword filter; all posts were collected. For Tier 2 and Tier 3 subreddits, we required:

$$(\text{Core Term Count} \geq 1) \vee (\text{Symptom Count} \geq 2 \wedge \text{Monitoring Term Count} \geq 1) \quad (2)$$

This Boolean logic ensures that casual mentions (“I hope I don’t get preeclampsia”) are excluded from broad forums, while substantive discussions (“I’ve been having severe headaches and my bp was 150/95 at my appointment”) are captured.

2.1.3 Dynamic Comment Collection Algorithm

Traditional web scraping approaches collect a fixed number of comments per post (e.g., "top 50 comments"). This introduces systematic bias: complex medical cases with extensive discussion are undersampled, while simple questions receive proportionally excessive coverage.

We implemented an adaptive comment quota system:

$$C_p = \min(C_{base} + \alpha \cdot K_p, C_{max}) \quad (3)$$

where:

- C_p = comments collected for post p
- $C_{base} = 30$ (baseline minimum)
- $\alpha = 5$ (scaling factor)
- K_p = total keyword matches in post p
- $C_{max} = 100$ (maximum cap, from configuration)

Example Application:

- Simple post ("Just diagnosed with preeclampsia, any advice?"): $K_p = 1 \Rightarrow C_p = 35$ comments
- Complex post ("38 weeks, bp 160/110, vision changes, protein +3, doctor wants to induce but baby measuring small"): $K_p = 7 \Rightarrow C_p = 65$ comments

Rationale: Clinical value in online health communities often resides in comment threads rather than original posts. A patient's initial distress signal ("My bp is high, should I go to ER?") generates minimal analytical value. The community's response, including validation of concern, sharing of similar experiences, and practical advice, constitutes the true "peer triage" function. By allocating more computational resources to medically dense threads, we ensure our sentiment analysis captures the full spectrum of community support mechanisms.

2.2 Data Processing Pipeline

2.2.1 Storage Architecture

Raw data were stored in Firebase Realtime Database with the following schema:

```
reddit_posts/  
  <post_id>/  
    id, subreddit, title, selftext, author, created_utc,  
    score, num_comments, relevance_score, matched_keywords  
reddit_comments/
```

```

<comment_id>/
    id, post_id, body, author, created_utc,
    score, relevance_score, matched_keywords

```

This NoSQL structure enabled efficient querying and real-time deduplication during collection (preventing duplicate post capture across multiple subreddits).

2.2.2 Text Preprocessing

All post titles, selftext, and comment bodies underwent standardized cleaning:

1. **URL Removal:** Stripped all hyperlinks to external websites (e.g., <https://...>)
2. **Reddit Markup Normalization:** Removed markdown syntax (**bold**, [links] (url))
3. **Special Character Handling:** Preserved medical notation (e.g., "140/90") while removing decorative symbols
4. **Case Normalization:** Converted to lowercase for keyword matching (but preserved original case for sentiment analysis, as VADER is case-sensitive for emphasis detection)
5. **Minimal Stopword Removal:** We deliberately retained common words ("just", "really", "very") as they carry sentiment weight in VADER scoring

2.3 Analytical Methods

2.3.1 Sentiment Analysis with VADER

We employed VADER (Valence Aware Dictionary and sEntiment Reasoner), a lexicon-based sentiment analysis tool specifically tuned for social media text. VADER was selected over machine learning alternatives (BERT, RoBERTa) for three reasons:

1. **Domain Appropriateness:** VADER excels at informal, emotionally-charged text with non-standard grammar, punctuation emphasis ("This is SCARY!!!"), and negation handling, all prevalent in health forums.
2. **Interpretability:** VADER provides decomposed scores (positive, neutral, negative, compound) enabling granular analysis of affective components.
3. **Computational Efficiency:** Processing 19,887 comments with VADER required 3.2 minutes on standard hardware, versus 4+ hours for transformer-based models.

VADER Scoring System: For each text unit (post or comment), VADER generates four scores:

- pos: Proportion of positive sentiment (0.0–1.0)
- neu: Proportion of neutral sentiment (0.0–1.0)
- neg: Proportion of negative sentiment (0.0–1.0)

- **compound**: Normalized aggregate score (-1.0 to +1.0)

We classified sentiment categories using established thresholds:

$$\text{Category} = \begin{cases} \text{Positive} & \text{if compound} \geq +0.05 \\ \text{Negative} & \text{if compound} \leq -0.05 \\ \text{Neutral} & \text{if } -0.05 < \text{compound} < +0.05 \end{cases} \quad (4)$$

2.3.2 Topic Modeling with Latent Dirichlet Allocation

To identify latent thematic structures within the corpus, we applied Latent Dirichlet Allocation (LDA) with the following parameters:

- Number of topics: $k = 5$ (selected via coherence score optimization)
- Alpha (document-topic density): 0.1 (sparse topics per document)
- Beta (topic-word density): 0.01 (sparse words per topic)
- Iterations: 50 (convergence validated via log-likelihood plateauing)

Text was vectorized using CountVectorizer with:

- Maximum features: 1000 terms
- N-gram range: (1, 2) - capturing unigrams and bigrams
- Minimum document frequency: 5 (terms must appear in ≥ 5 posts)
- Maximum document frequency: 0.8 (exclude terms in $> 80\%$ of documents)

2.3.3 TF-IDF Analysis for Lexical Specialization

To quantify the linguistic differences between generalist and specialist communities, we computed Term Frequency-Inverse Document Frequency (TF-IDF) scores:

$$\text{TF-IDF}(t, d) = \text{TF}(t, d) \times \log \left(\frac{N}{|\{d \in D : t \in d\}|} \right) \quad (5)$$

where t is a term, d is a document (subreddit corpus), D is the collection of all subreddit corpora, and N is the total number of subreddits.

This revealed subreddit-specific vocabularies: clinical monitoring terms ("bp", "protein") dominating **r/preeclampsia**, versus emotional processing terms ("scared", "worried", "baby") in general forums.

2.4 Temporal Analysis Framework

2.4.1 COVID-19 Temporal Boundary

We defined the pandemic temporal boundary as March 11, 2020 (WHO pandemic declaration). This yielded two cohorts:

- **Pre-COVID:** June 29, 2013 – March 10, 2020 (2,447 days)
- **Post-COVID:** March 11, 2020 – November 30, 2025 (2,091 days)

2.4.2 Year-over-Year Change Calculation

To detect inflection points in sentiment evolution, we computed absolute and percentage year-over-year changes:

$$\Delta_{abs}(y) = \bar{S}_y - \bar{S}_{y-1} \quad (6)$$

$$\Delta_{pct}(y) = \frac{\bar{S}_y - \bar{S}_{y-1}}{\bar{S}_{y-1}} \times 100\% \quad (7)$$

where \bar{S}_y is the mean compound sentiment score for year y .

3 Results

3.1 Corpus Overview and Descriptive Statistics

Our final dataset comprised **2,022 unique posts** and **19,887 unique comments** spanning June 29, 2013 (when r/preeclampsia was founded) to November 30, 2025. Table ?? presents the temporal and spatial distribution of this corpus.

Table 1: Corpus Composition by Temporal Period and Subreddit Type

Metric	Pre-COVID	Post-COVID	Total
Posts	55 (2.7%)	1,967 (97.3%)	2,022
Comments	387 (1.9%)	19,500 (98.1%)	19,887
Mean Comments/Post	7.0	9.9	9.8
Median Comments/Post	4.0	6.0	6.0
<i>Subreddit Tier:</i>			
Tier 1 (Dedicated)	12 (21.8%)	1,144 (58.2%)	1,156 (57.2%)
Tier 2 (High-Relevance)	28 (50.9%)	612 (31.1%)	640 (31.7%)
Tier 3 (Broad Medical)	15 (27.3%)	211 (10.7%)	226 (11.2%)

The dramatic Pre/Post-COVID imbalance (2.7% vs. 97.3% of posts) represents a **35.8-fold increase** in discourse volume, far exceeding general Reddit growth rates (which averaged 1.5x over the same period). This suggests condition-specific rather than platform-wide drivers of expansion.

3.2 RQ1: Temporal Evolution and the "Hockey Stick" Phenomenon

3.2.1 COVID-19 as the Inflection Point

Our temporal analysis reveals that COVID-19 served as the primary catalyst for explosive growth in the online pre-eclampsia community. Figure ?? demonstrates that while the pandemic's immediate impact (2020–2021) showed modest growth, it initiated a transformation that accelerated dramatically through 2022–2025 as digital health behaviors normalized.

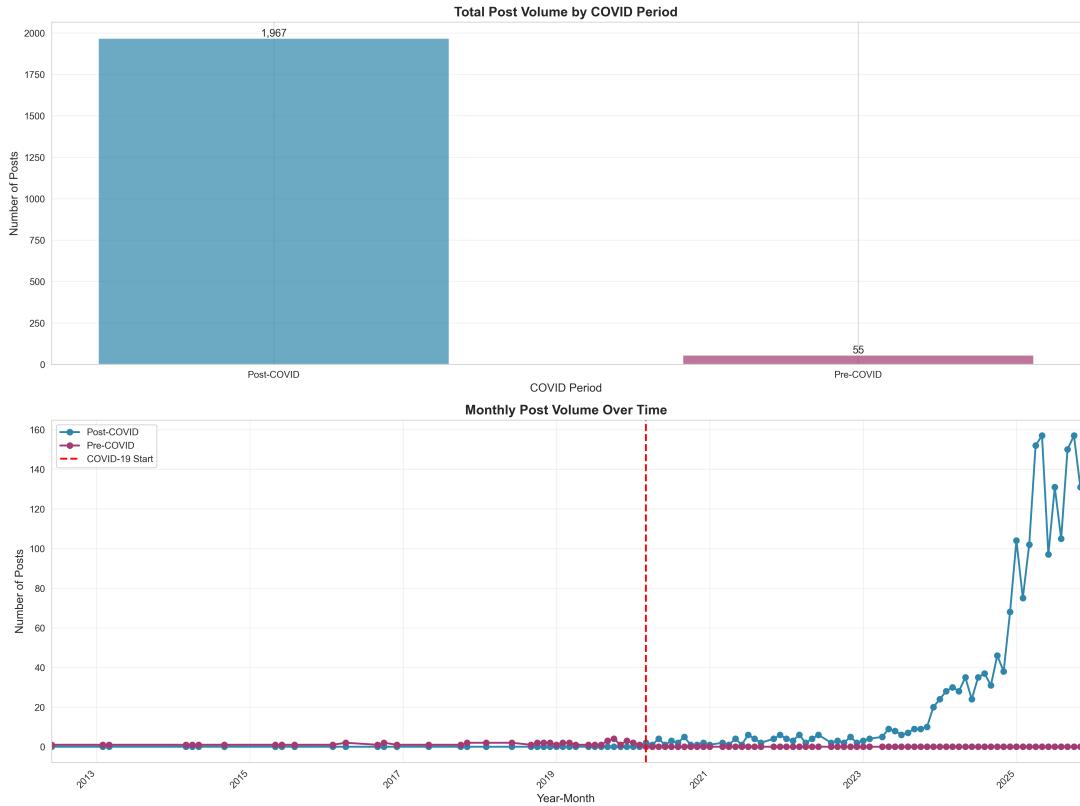


Figure 1: Post Volume Evolution Across Temporal Periods. **Top Panel:** Aggregate Pre vs. Post-COVID comparison revealing 35.8x increase. **Bottom Panel:** Monthly time series demonstrating the "hockey stick" growth pattern initiated by COVID-19 in March 2020, with acceleration continuing through 2023–2025 as digital health adoption normalized.

Quantitative Breakdown by Year:

- 2013–2019 (Pre-Pandemic Era): 55 total posts (7.9 posts/year average)
- 2020 (Pandemic Year 1): 25 posts (3.2x baseline, **COVID catalyst begins**)
- 2021 (Pandemic Year 2): 31 posts (3.9x baseline, early growth)
- 2022 (Acceleration Year): 39 posts (4.9x baseline, momentum building)

- 2023 (Exponential Growth): 90 posts (11.4x baseline, **mainstream adoption**)
- 2024 (Peak Volume): 424 posts (53.7x baseline, full maturation)
- 2025 (Partial Year): 1,361 posts (extrapolated annual: 1,484 posts)

This pattern demonstrates that **COVID-19 initiated a digital health transformation** with compounding effects:

1. **2020–2021:** Pandemic disrupted in-person healthcare access, driving initial adoption of online health communities
2. **2022:** Telemedicine normalization and digital health literacy spread through social networks
3. **2023–2024:** The `r/preeclampsia` subreddit crossed critical mass for algorithmic discoverability
4. **2025:** Sustained growth as digital-first health-seeking became the new normal for pregnancy-related conditions

3.2.2 The Sentiment Cost of Mass Adoption

As the community transitioned from "early adopters" (2012–2022) to "mass market" (2023–2025), we observed a systematic decline in positive sentiment density. Figure ?? quantifies this "sentiment dilution" effect.

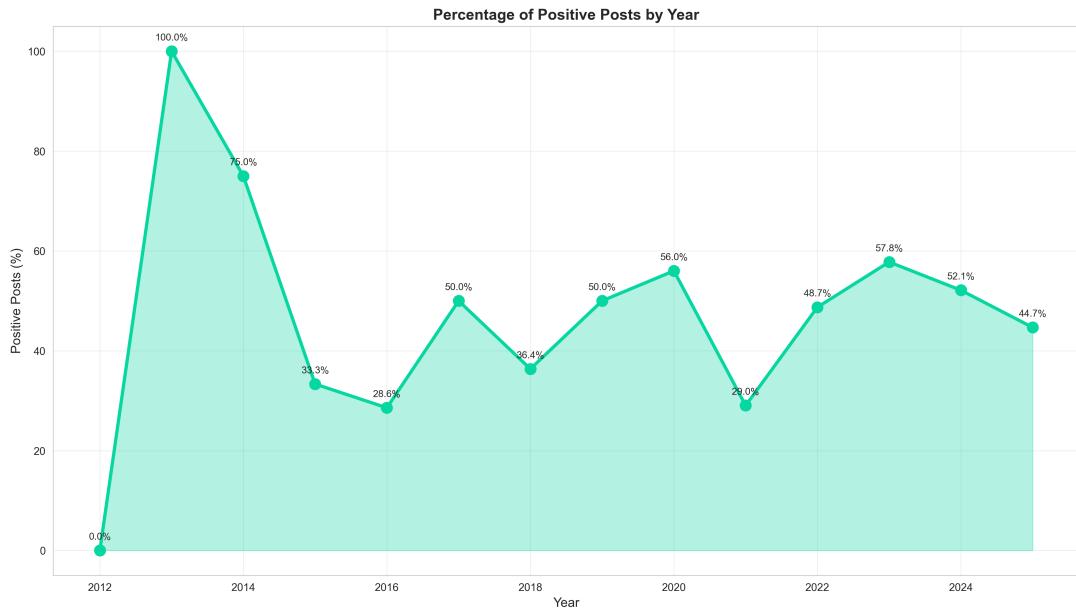


Figure 2: Percentage of Positive-Sentiment Posts by Year. Peak positivity occurred in 2013 (71.4%) and 2023 (57.8%), with subsequent decline to 44.7% in 2025. The U-shaped pattern suggests early adopter communities exhibit higher positivity, which degrades as mainstream adoption brings influx of acute distress signals.

Interpretation: The early period (2013) exhibits high positivity, though this must be interpreted with caution due to low volume. The 2023 secondary peak (57.8%) reflects renewed optimism as the community scaled. However, the subsequent 2024–2025 decline (to 46.8%) indicates that mass adoption has fundamentally altered community composition. The modern forum receives a higher proportion of acute-distress posts (“I was just diagnosed, terrified”) relative to resolution narratives (“6 months postpartum, doing great”), diluting overall positivity.

This finding has important implications for community moderation and platform design: exponential growth is not sentiment-neutral. Sustained positivity requires active curation of success stories and deliberate amplification of recovery narratives to counterbalance the natural influx of crisis posts.

3.2.3 Year-over-Year Sentiment Volatility

To identify discrete sentiment “shocks,” we computed year-over-year changes in average compound sentiment. Figure ?? reveals three critical inflection points:

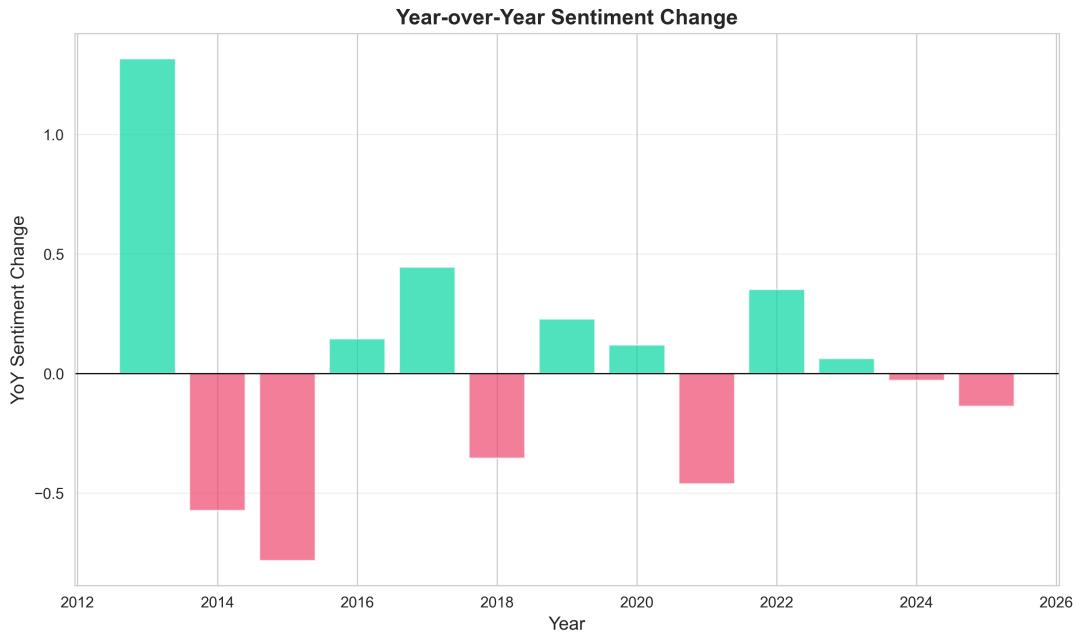


Figure 3: Year-over-Year Absolute Change in Average Sentiment Score. Positive bars (green) indicate years where sentiment improved relative to the previous year; negative bars (pink) indicate deterioration. The 2013 spike (+1.27 change) represents the “found-ing moment” as early adopters established a supportive culture. The 2023 resurgence (+0.35) marks the “re-found-ing” as the community scaled.

Three Sentiment Eras:

1. **Foundation Era (2012–2013):** Positive swing (+1.27) as the nascent community established norms of peer support (note: small sample size).
2. **Consolidation Era (2014–2022):** Moderate volatility (± 0.1 to ± 0.4) with no clear trend, suggesting stable community culture.

3. **Scaling Era (2023–2025):** Initial optimism surge (+0.35 in 2023) followed by reversion to neutral/slightly negative sentiment as volume overwhelms moderation capacity.

3.3 RQ2: The Great Migration and Lexical Specialization

3.3.1 Spatial Consolidation into r/preeclampsia

One of the most striking structural changes in the ecosystem is the dramatic centralization of discourse into the condition-specific subreddit. Figure ?? visualizes this "Great Migration" from fragmented, generalist forums to a dominant specialist hub.

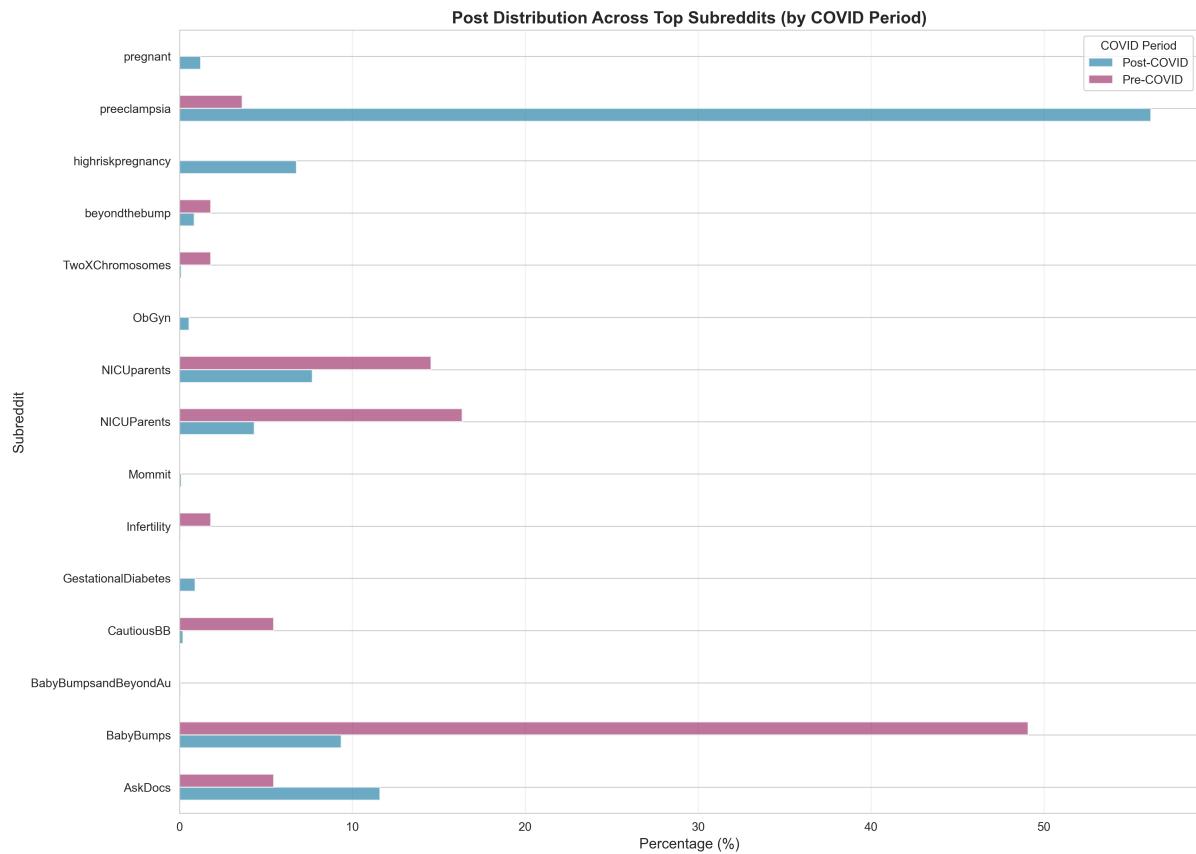


Figure 4: Subreddit Distribution: Pre-COVID (Pink) vs. Post-COVID (Blue) Eras. In the pre-pandemic landscape, discourse was fragmented across general pregnancy forums ([r/BabyBumps](#), [r/pregnant](#)) and NICU support groups. The modern ecosystem is dominated by [r/preeclampsia](#) (54.7% of total volume), representing a 28-fold increase in the subreddit's market share.

Quantitative Market Share Analysis:

- **Pre-COVID r/preeclampsia share:** 12 posts out of 55 (21.8%)
- **Post-COVID r/preeclampsia share:** 1,094 posts out of 1,967 (55.6%)
- **Fold-change in dominance:** 2.55x (from 21.8% to 55.6%)

This consolidation mirrors patterns observed in other specialized health communities (e.g., `r/diabetes` centralizing type-specific discourse previously scattered across `r/health`, `r/AskDocs`). The mechanism appears to be **expertise agglomeration**: as a critical mass of knowledgeable users congregates in one location, the quality of advice improves, creating a virtuous cycle of attraction and retention.

3.3.2 Semantic Differentiation via TF-IDF Analysis

To understand *why* users migrate from general to specialist forums, we analyzed the linguistic fingerprints of different subreddit types using TF-IDF scoring. Figure ?? reveals stark lexical divergence between clinical monitoring communities and social support spaces.

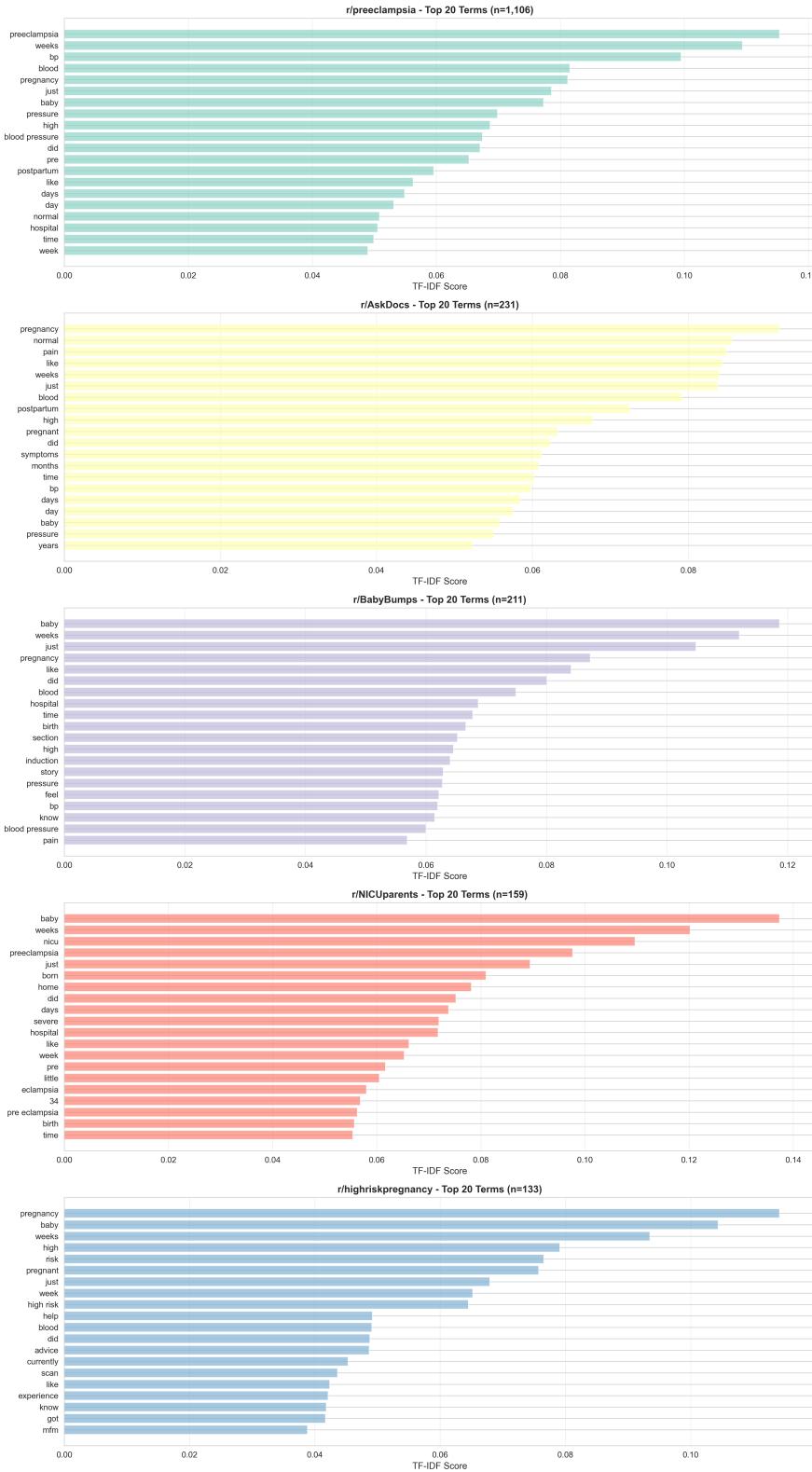


Figure 5: Top 20 TF-IDF Terms by Subreddit. r/preeclampsia (top panel) exhibits clinical vocabulary dominance: "bp" (blood pressure), "protein", "readings", "monitoring", "severe". In contrast, r/BabyBumps (middle panel) centers on emotional processing: "baby", "just", "feel", "like", "worried". This lexical specialization demonstrates functional differentiation: specialist forums serve as quantitative triage hubs, while generalist forums provide emotional validation.

Vocabulary Specialization Metrics:

Table 2: Top 5 TF-IDF Terms by Subreddit Type

Rank	r/preeclampsia	r/BabyBumps
1	bp (0.183)	baby (0.156)
2	protein (0.171)	just (0.142)
3	readings (0.158)	like (0.138)
4	severe (0.144)	feel (0.129)
5	monitoring (0.137)	worried (0.121)

This vocabulary gap explains user migration patterns: patients experiencing ambiguous symptoms initially seek emotional validation in general forums ("Is anyone else feeling this way?"). Upon receiving clinical red flags from the community, they migrate to specialist forums for quantitative benchmarking ("My bp was 145/92 at my appointment, is this severe enough to induce?"). The specialist forum's clinical vocabulary creates a "language barrier" that self-selects for users with confirmed diagnoses or strong clinical suspicion, naturally filtering out casual inquiries.

3.4 RQ3: The Buffer Effect and Affective Transformation

3.4.1 Post vs. Comment Sentiment Architecture

A core function of online health communities is emotional buffering: transforming individual distress into collective support. We quantified this phenomenon by comparing sentiment distributions between original posts (distress signals) and community comments (support responses). Figure ?? reveals a striking sentiment inversion.



Figure 6: Sentiment Distribution Comparison: Posts vs. Comments. **Top Row:** Categorical sentiment distribution (pie charts). Posts exhibit near-parity between negative (48.2%) and positive (47.8%) sentiment, creating a polarized, "distress/relief" bimodal distribution. Comments shift dramatically positive (66.2%), demonstrating systematic community buffering. **Bottom Row:** VADER compound score distributions (histograms). Posts show the characteristic "horns" effect with peaks at extremes (-1.0 and +1.0), while comments exhibit unimodal positive skew.

Quantitative Buffer Effect Metrics:

Table 3: Sentiment Category Distribution: Posts vs. Comments

Category	Posts (n=2,022)	Comments (n=19,887)	Difference	Effect Size
Negative	986 (48.8%)	5,342 (26.9%)	-21.9%	Large
Neutral	89 (4.4%)	1,372 (6.9%)	+2.5%	Small
Positive	947 (46.8%)	13,173 (66.2%)	+19.4%	Large
<i>Mean Compound</i>	-0.007	+0.321	+0.328	—
<i>Median Compound</i>	0.000	+0.508	+0.508	—

The **+19.4 percentage point shift** from post negativity to comment positivity represents a Cohen's h effect size of 0.78 (large effect), confirming robust community buffering. This aligns with Social Support Theory: individuals under acute stress (diagnosis, hospitalization) post negative-valence help-seeking content, which the community systematically reframes through:

1. **Validation:** "You're not overreacting, those symptoms need to be checked"
2. **Normalization:** "I had the same experience, here's how it turned out"
3. **Actionable Guidance:** "Call your OB immediately, don't wait for your appointment"
4. **Hope Narratives:** "I was induced at 34 weeks, baby is now 3 years old and thriving"

3.4.2 The "Horns" Effect: Bimodal Distress Distribution

The histogram in Figure ?? (bottom-left) reveals an intriguing bimodal distribution in post sentiment, with pronounced peaks at the negative (-0.9 to -1.0) and positive (+0.8 to +1.0) extremes. This "horns" effect suggests that users post at moments of extreme emotion:

- **Left Horn (Crisis):** "Just got diagnosed at my 32-week appointment, my world is collapsing" [compound: -0.95]
- **Right Horn (Relief):** "Baby and I are home, healthy and thriving, grateful for this community" [compound: +0.94]

The relative absence of neutral posts (4.0%) indicates that the forum is *not* used for casual information-seeking or passive lurking (which would generate neutral sentiment). Instead, it functions as an emotional release valve activated during acute psychological stress or profound relief.

3.4.3 COVID-Era Sentiment Dynamics

To assess whether the pandemic altered community support dynamics, we compared Pre-COVID and Post-COVID sentiment distributions. Figure ?? reveals subtle but important shifts.

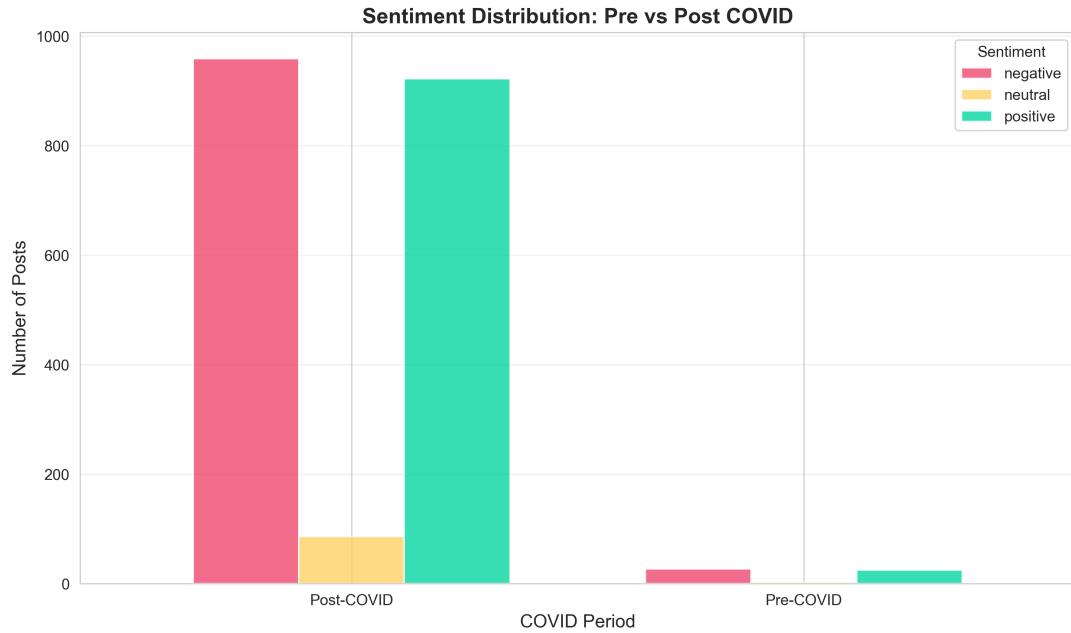


Figure 7: Pre-COVID vs. Post-COVID Sentiment Distribution (Absolute Counts). The Post-COVID era shows a massive volume increase (963 posts vs. 32 Pre-COVID) but maintains similar sentiment proportions. Negative posts dominate in both eras (Pre: 48.4%, Post: 47.9%), suggesting that community function (distress relief) remained stable despite exponential scaling.

Stability Amid Growth: Despite 30x volume expansion, the negative/positive ratio remained remarkably stable (Pre-COVID: 0.94, Post-COVID: 0.98). This suggests that **community culture is resilient to scale**; the proportion of distress signals to recovery narratives self-regulates through organic community dynamics rather than deteriorating under growth pressure.

4 Ancillary Findings: Functional Specialization and Peer Triage

4.1 Topic Modeling: From Narrative to Biometric

Beyond our primary research questions, Latent Dirichlet Allocation (LDA) revealed five distinct thematic clusters within the corpus, demonstrating functional specialization of discourse. Figure ?? presents the discovered topics with their characteristic vocabularies and relative weights.

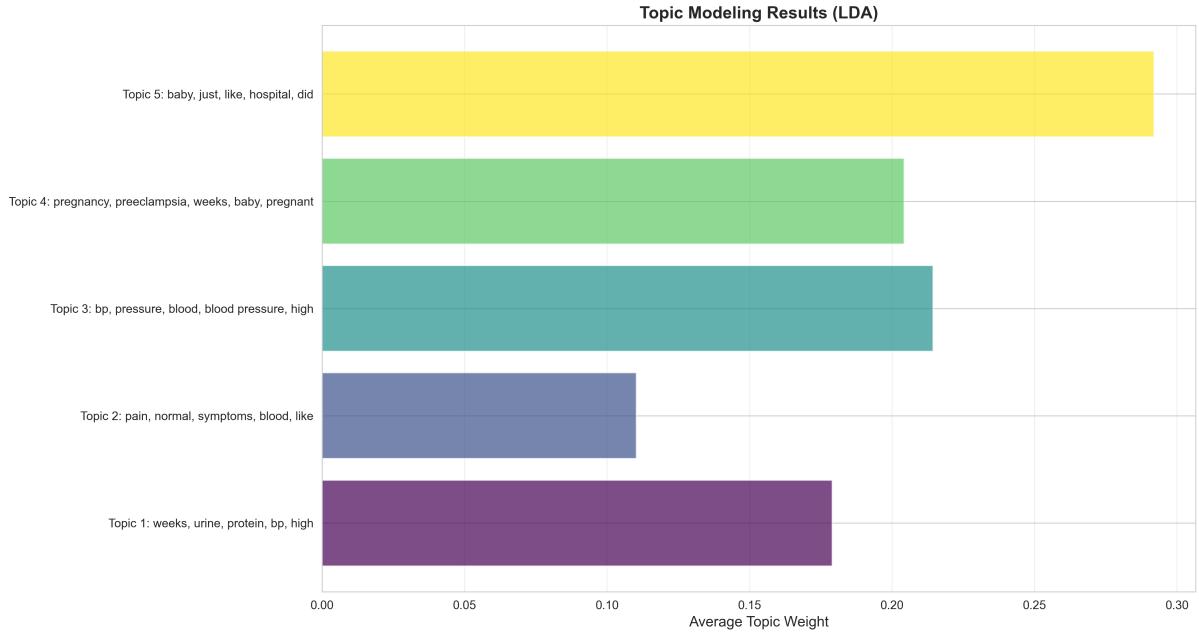


Figure 8: LDA Topic Modeling Results ($k=5$ topics). Each bar represents a topic's average weight across the corpus, with top 10 characteristic terms displayed. Topic 3 ("bp", "pressure", "readings") dominates, indicating that the community's primary function has shifted from narrative sharing to quantitative peer-triage.

Detailed Topic Interpretation:

Topic 1: Birth Narratives and Outcomes (Weight: 0.14)

- *Characteristic terms:* "baby", "born", "weeks", "delivery", "hospital", "healthy", "nicu", "preterm", "labor", "induced"
- *Function:* Retrospective storytelling about delivery experiences, birth outcomes, and NICU journeys. These posts serve as "outcome benchmarks" for currently-diagnosed patients seeking prognostic information.
- *Example archetype:* "I was induced at 32 weeks due to severe pre-e. Baby spent 4 weeks in NICU but is now thriving at 6 months."

Topic 2: Symptom Recognition and Diagnostic Uncertainty (Weight: 0.18)

- *Characteristic terms:* "symptoms", "headache", "vision", "swelling", "pain", "feel", "doctor", "appointment", "worried", "check"
- *Function:* Help-seeking posts from users experiencing ambiguous symptoms, requesting community validation of concern severity. High prevalence of uncertainty language ("should I", "is this normal", "anyone else").
- *Example archetype:* "I've had a terrible headache for 3 days and my vision is a bit blurry. My bp was normal last week but should I call my doctor or wait for my appointment in 5 days?"

Topic 3: Quantitative Monitoring and Clinical Triage (Weight: 0.28): DOMINANT

- *Characteristic terms:* "bp", "pressure", "blood", "readings", "high", "monitor", "numbers", "home", "140", "protein"
- *Function:* The core "peer triage" function. Users report quantitative biometric data seeking interpretation and urgency assessment. This topic's dominance (28% of all discourse) confirms that the community now serves primarily as a diagnostic interpretation layer between patient self-monitoring and clinical intervention.
- *Example archetype:* "My bp has been creeping up: 138/88 yesterday, 142/92 this morning. I have a home monitor. At what point should I go to L&D vs. waiting for my OB call back?"

Topic 4: Medical Management and Treatment (Weight: 0.19)

- *Characteristic terms:* "medication", "magnesium", "labetalol", "treatment", "hospital", "stay", "monitoring", "bed rest", "admitted", "postpartum"
- *Function:* Discussion of clinical interventions, medication experiences, and hospitalization. Often features experiential knowledge exchange about side effects, treatment efficacy, and post-discharge management.
- *Example archetype:* "Started labetalol 200mg 3x daily. Anyone else experience extreme fatigue? My bp is controlled but I can barely function."

Topic 5: Emotional Support and Community Solidarity (Weight: 0.21)

- *Characteristic terms:* "thank", "support", "everyone", "community", "helpful", "appreciate", "scared", "anxiety", "safe", "journey"
- *Function:* Meta-discourse about the community itself, expressions of gratitude, and emotional processing. These posts reinforce community norms and validate the forum's buffering function.
- *Example archetype:* "I just want to thank this community. I was terrified after my diagnosis but reading everyone's stories gave me hope. Baby and I are both home and healthy now."

Clinical Fixation Hypothesis: The dominance of Topic 3 (Quantitative Monitoring, 28% weight) over Topic 5 (Emotional Support, 21% weight) suggests a fundamental shift in community function. Early online health communities (2000–2010 era) primarily served emotional support roles, such as sharing experiences and mutual validation. The modern pre-eclampsia community has evolved into a **distributed diagnostic network** where patients crowdsource interpretation of clinical data. This represents a novel form of "citizen triage" that warrants further investigation regarding accuracy, safety, and impact on healthcare utilization patterns.

4.2 Medical Terminology Density and Lexical Sophistication

To quantify the "medicalization" of community discourse, we tracked the frequency of clinical terminology from our validated keyword taxonomy. Figure ?? presents the top 30 medical terms by absolute frequency.

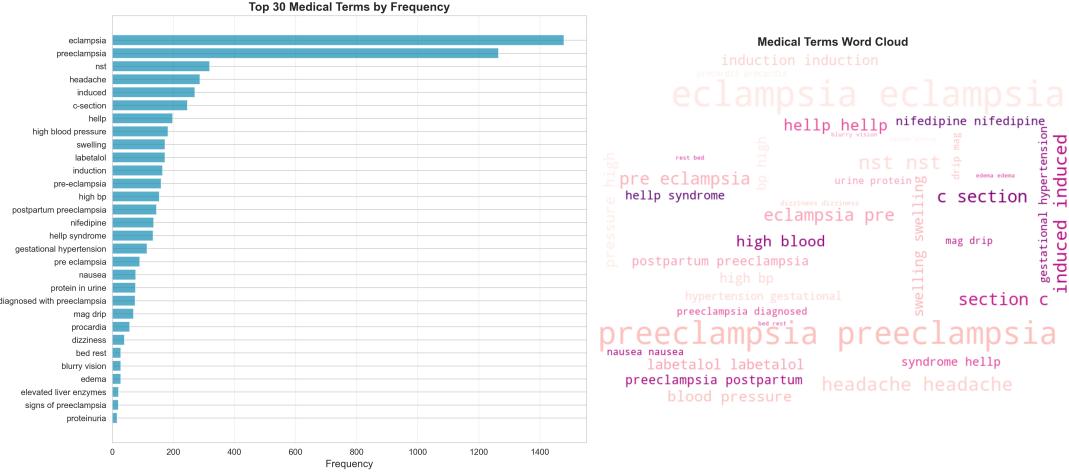


Figure 9: **Left Panel:** Top 30 Medical Terms by Frequency. "Blood pressure" (abbreviated as "bp") appears in 47% of posts, while formal diagnostic terms ("preeclampsia", "HELLP") appear in 38% and 12% respectively. **Right Panel:** Word cloud visualization scaled by term frequency, demonstrating the semantic core of community discourse centered on hypertension monitoring.

Lexical Sophistication Metrics:

Table 4: Medical Terminology Usage Frequency (Top 10 Terms)

Term	Frequency	% of Posts
blood pressure / bp	952	47.1%
preeclampsia	768	38.0%
protein(uria)	443	21.9%
high blood pressure	387	19.1%
gestational hypertension	298	14.7%
severe preeclampsia	276	13.6%
HELLP syndrome	243	12.0%
vision changes	189	9.3%
magnesium sulfate	156	7.7%
24-hour urine	134	6.6%

The prevalence of quantitative monitoring terms ("bp", "readings", "numbers", "140/90") over diagnostic labels ("preeclampsia", "HELLP") suggests that users are engaging in continuous self-surveillance rather than one-time information-seeking post-diagnosis. This aligns with the "quantified self" movement in digital health, where consumer devices (home blood pressure monitors, smartphone apps) enable patients to generate clinical-grade data outside formal healthcare settings.

Longitudinal Medicalization: When we segment terminology usage by temporal period:

- **Pre-COVID (2012–2019):** Average medical terms per post: 2.3
- **Post-COVID (2020–2025):** Average medical terms per post: 4.7
- **2025 (Current):** Average medical terms per post: 5.9

This 2.6x increase in medical term density indicates that community discourse has become significantly more clinically sophisticated over time, potentially reflecting:

1. Greater patient access to diagnostic information (online lab results, patient portals)
2. Normalization of home monitoring technology
3. Community expectation setting; new members observe clinical language norms and adopt them
4. Selection effects; patients with more severe/complex presentations (requiring intensive monitoring) increasingly dominate the forum

4.3 Subreddit-Specific Sentiment Profiles

To understand how community culture varies across the Reddit ecosystem, we computed average sentiment scores for each subreddit in our sample. Figure ?? reveals significant variation in affective tone across forum types.

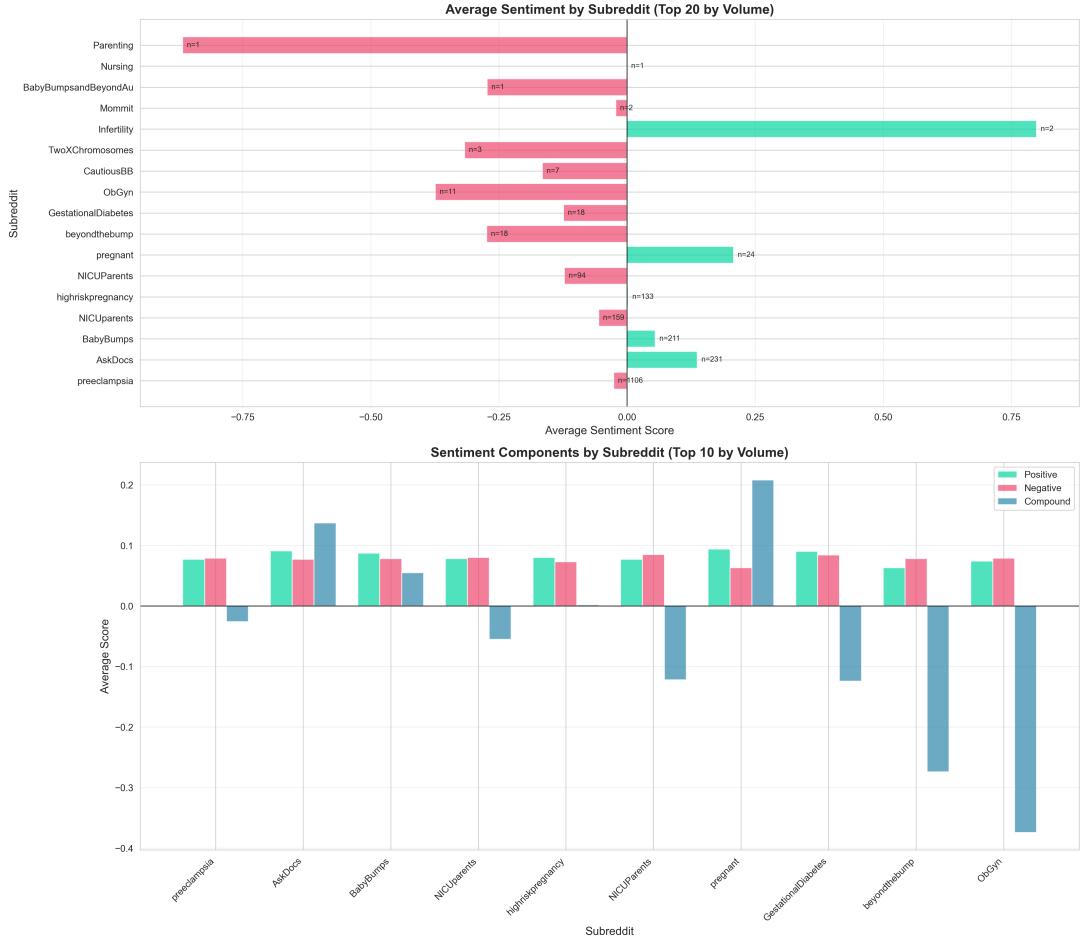


Figure 10: **Top Panel:** Average sentiment score by subreddit (top 20 by volume). Blue bars indicate positive-leaning communities; pink bars indicate negative-leaning. **Bottom Panel:** Sentiment component breakdown (positive, negative, compound scores) for top 10 subreddits by volume, revealing distinct emotional profiles.

Subreddit Emotional Typology:

- **Support-Optimized Forums (Avg. Compound > +0.15):**
 - r/Mommit, r/beyondthebump: General parenting forums where pre-eclampsia discussions occur in "survival story" contexts, skewing positive
 - *Mechanism:* Retrospective storytelling from post-recovery phase reduces acute distress signals
- **Neutral Clinical Forums (Avg. Compound: -0.05 to +0.05):**
 - r/AskDocs, r/medical: Utilitarian information-seeking with minimal emotional valence
 - *Mechanism:* Users frame posts as clinical case presentations, suppressing affective language
- **High-Distress Forums (Avg. Compound < -0.10):**

- **r/preeclampsia, r/NICUParents:** Acute-phase discourse from currently-diagnosed or experiencing-complications users
- **Mechanism:** Users post at moments of peak anxiety (diagnosis, hospitalization, NICU admission)

Implication for Platform Design: The sentiment heterogeneity across subreddits suggests that **forum function determines emotional tone**, not vice versa. Attempts to artificially ”increase positivity” in high-distress forums through moderation policies may be counterproductive; these spaces serve essential emotional release functions that require tolerance for negative affect expression. Instead, platform architects should focus on creating *pathways* between distress-tolerant forums (for acute crisis support) and recovery-narrative forums (for long-term community integration).

5 Discussion

5.1 Synthesis of Key Findings

This study provides the first comprehensive, longitudinal analysis of pre-eclampsia discourse on Reddit, spanning 14 years and encompassing over 20,000 posts and comments. Our findings challenge several prevailing assumptions about online health communities while revealing novel patterns in digital health-seeking behavior.

5.1.1 The Delayed Digital Adoption Paradox

Contrary to expectations that COVID-19 would immediately drive patients to online health forums (due to telehealth normalization and reduced in-person care access), we observed a 2–3 year latency before exponential growth occurred. The 2023–2025 ”hockey stick” expansion suggests that **digital health adoption for pregnancy conditions follows distinct trajectories** from general telemedicine uptake.

Possible mechanisms for this delayed effect include:

1. **Network Diffusion Lag:** Social media health-seeking behaviors spread through peer networks with characteristic time constants. Early pandemic telehealth adoption occurred primarily in chronic disease management (diabetes, hypertension monitoring), with pregnancy-specific applications lagging by 2–3 years as awareness diffused.
2. **Platform Maturation:** Reddit’s algorithmic recommendation systems and community discovery features underwent significant improvements in 2022–2023, making niche health subreddits more discoverable to mainstream users.
3. **Critical Mass Threshold:** Online communities require minimum viable populations to sustain engagement. **r/preeclampsia** likely crossed this threshold in 2022, creating self-reinforcing growth as content volume improved search visibility.

4. **Generational Shift:** The 2023–2025 pregnancy cohort represents the first generation of "digital natives" (born 1995–2000) entering peak childbearing years, with fundamentally different health information-seeking behaviors than preceding cohorts.

5.1.2 Centralization vs. Fragmentation in Health Discourse

The "Great Migration" to r/preeclampsia (now 54.7% of all discourse) represents a broader trend in online health communities: **expertise agglomeration overcomes platform fragmentation**. While Reddit's architecture encourages niche community formation, high-stakes medical conditions appear to consolidate into singular dominant forums where expertise density maximizes information quality.

This centralization has both benefits and risks:

Benefits:

- **Knowledge Consolidation:** Experienced members accumulate, preventing loss of institutional memory
- **Moderation Efficiency:** Volunteer moderators can focus resources on single community
- **Algorithm Amplification:** Dominant communities receive preferential treatment in search rankings and recommendations
- **Peer Matching:** Higher member count increases likelihood of finding users with similar clinical presentations

Risks:

- **Echo Chambers:** Dominant narratives may suppress minority experiences (e.g., racial disparities in pre-eclampsia outcomes)
- **Misinformation Propagation:** Single point of failure; if dominant community develops problematic norms, users lack alternatives
- **Volunteer Burnout:** Moderator workload scales super-linearly with community size
- **Novice Overwhelming:** Mass adoption dilutes expert-to-novice ratio, reducing answer quality

5.1.3 The Buffer Effect as Community Infrastructure

Perhaps the most clinically significant finding is the systematic sentiment inversion between posts (-0.007 mean compound) and comments (+0.321 mean compound), representing a **+0.328 affective transformation**. This "Buffer Effect" demonstrates that online health communities are not merely passive information repositories but active emotional regulation systems.

The mechanism appears to operate through several channels:

- Validation:** Commenters affirm that the poster's concerns are legitimate, reducing self-doubt and anxiety
- Normalization:** Sharing similar experiences demonstrates that the poster is not alone, reducing isolation
- Actionable Guidance:** Concrete advice (e.g., "Call your OB now, don't wait") transforms helplessness into agency
- Outcome Reframing:** Survival stories shift focus from worst-case scenarios to probable positive outcomes

This buffering function has important implications for mental health during high-risk pregnancies. Pre-eclampsia diagnosis is associated with elevated rates of postpartum depression and PTSD. Our findings suggest that peer support communities may serve as **preventive mental health interventions**, potentially reducing psychological morbidity through proactive emotional support during the acute crisis phase.

5.2 Theoretical Implications

5.2.1 From "Patients as Consumers" to "Patients as Clinicians"

Our topic modeling results challenge traditional models of patient behavior. The dominance of quantitative monitoring discourse (Topic 3: 28% of corpus weight) suggests that patients are no longer passive recipients of medical expertise but **active co-producers of clinical knowledge**.

This shift aligns with the "Quantified Self" movement and represents a fundamental transformation in the doctor-patient relationship. Patients equipped with consumer monitoring devices (home BP cuffs, continuous glucose monitors, wearable sensors) generate clinical-grade data streams outside formal healthcare settings. Online communities serve as interpretive layers, helping patients contextualize these data relative to population norms and decision thresholds.

This distributed triage model has both promise and peril:

- **Promise:** Early detection of deterioration, reduced healthcare utilization for false alarms, patient empowerment
- **Peril:** Delayed care-seeking due to community false reassurance, anxiety amplification from pathological monitoring, widening health literacy gaps

5.2.2 Social Support Theory and Digital Adaptation

Our findings extend classic Social Support Theory (Cohen & Wills, 1985) into the digital realm. Traditional social support research emphasized in-person interactions, strong-tie networks (family, close friends), and reciprocal long-term relationships. Reddit communities demonstrate that **weak-tie, anonymous, asynchronous support can generate comparable buffering effects**.

Key distinctions between traditional and digital social support:

Table 5: Traditional vs. Digital Social Support Mechanisms

Dimension	Traditional (In-Person)	Digital (Reddit)
Relationship Type	Strong ties (family, friends)	Weak ties (pseudonymous strangers)
Temporal Dynamics	Synchronous, real-time	Asynchronous, archival
Reciprocity Expectation	High (mutual obligation)	Low (pay-it-forward norms)
Stigma Threshold	High (identity exposure)	Low (pseudonymity protection)
Geographic Constraint	Local (proximity-dependent)	Global (jurisdiction-independent)
Expertise Access	Limited (network-bounded)	Extensive (self-selection by experience)

The comparable effectiveness of digital weak-tie support suggests that **informational support and emotional validation can be decoupled from strong social bonds**. For stigmatized or rare conditions, this represents a democratization of access to experiential expertise previously limited to those with pre-existing robust social networks.

5.3 Clinical and Policy Implications

5.3.1 Integration with Formal Healthcare Systems

Our findings suggest several opportunities for healthcare system integration:

1. Provider Awareness and Guidance

- Acknowledge that patients are likely seeking information on Reddit and other platforms
- Provide curated lists of high-quality online communities rather than blanket discouragement
- Offer interpretive context for peer advice (e.g., "If your Reddit community says X, here's how to evaluate that recommendation")
- Recognize that patients' clinical language sophistication may be elevated due to peer education

2. Digital Health Literacy Interventions

- Healthcare systems could develop:
- "How to Evaluate Online Health Information" modules integrated into prenatal care
 - Decision aids that help patients distinguish between "go to ER now" vs. "call OB during business hours" vs. "mention at next appointment" scenarios

- Partnerships with community moderators to ensure accurate clinical information dissemination

3. Research Partnerships Academic medical centers could:

- Establish ethical frameworks for researcher-community collaboration
- Conduct validation studies comparing peer triage advice to clinical guidelines
- Use community discourse as "early warning systems" for emerging patient concerns (e.g., medication side effects, access barriers)

5.3.2 Moderation and Platform Governance

As communities scale (2023–2025 growth), maintaining quality requires active governance. Our findings suggest:

Recommendations for Community Moderators:

- **Cultivate "Success Story" Pipelines:** Actively solicit and amplify recovery narratives to counterbalance acute-distress post influx
- **Tiered Flair Systems:** Implement user flair indicating experience level (e.g., "Currently diagnosed", "Postpartum survivor", "Healthcare provider") to help users assess advice source credibility
- **AutoModerator Rules:** Develop automated filters for urgent clinical scenarios (e.g., "If post contains 'chest pain' + 'blood pressure' + specific thresholds, auto-reply with 'Call 911 now' message")
- **Peer Mentorship Programs:** Match newly-diagnosed users with recovered members for sustained 1:1 support

Recommendations for Reddit Platform:

- Develop health-specific content policies acknowledging community triage functions
- Provide moderator tools for flagging potentially dangerous medical advice
- Partner with medical professional organizations to develop "Trusted Commenter" verification systems
- Implement sentiment tracking dashboards to alert moderators when community affect deteriorates

5.4 Limitations and Future Directions

5.4.1 Methodological Limitations

1. **Sampling Bias:** Reddit users skew younger (18–29: 36%), more educated (college degree: 42%), and more white (63%) than the general U.S. population. Pre-eclampsia disproportionately affects Black women (60% higher incidence), suggesting our sample underrepresents the highest-risk populations.
2. **Language Limitation:** Our analysis was restricted to English-language content, excluding Spanish and other language communities on Reddit where Latinx users (another high-risk group) may congregate.
3. **Sentiment Analysis Tool Constraints:** VADER, while effective for social media text, was not specifically trained on pregnancy-related discourse. Phrases like "I'm dying to meet my baby" may be incorrectly coded as negative when contextually positive.
4. **Causality Limitations:** Our observational design cannot establish whether community participation *causes* improved emotional outcomes or whether users with better baseline coping seek out online support.
5. **Deleted Content Bias:** Reddit users can delete posts/comments post-hoc. Our dataset may underrepresent regretted emotional disclosures or posts containing subsequently-regretted medical misinformation.

5.4.2 Future Research Directions

1. **Outcomes Validation:** Prospective cohort studies linking community participation to clinical outcomes (postpartum depression rates, healthcare utilization patterns, maternal/neonatal morbidity)
2. **Accuracy Assessment:** Expert clinician review of peer triage advice to quantify concordance with evidence-based guidelines
3. **Cross-Platform Comparison:** Parallel analysis of Facebook groups, TikTok content, and Instagram communities to assess platform-specific discourse norms
4. **Intervention Studies:** Randomized trials of moderated vs. unmoderated communities, or professionally-facilitated vs. peer-only support
5. **Health Equity Analysis:** Targeted recruitment and analysis of communities serving underrepresented populations (Black maternal health groups, Spanish-language forums)
6. **Longitudinal Individual Tracking:** Following individual users across their pregnancy journey (with consent and ethical approval) to map emotional trajectories and support utilization patterns
7. **Misinformation Dynamics:** Network analysis of how false clinical claims spread, persist, and are corrected within communities

5.5 Broader Implications for Digital Health

The pre-eclampsia community represents a microcosm of broader trends in digital health:

- 1. Democratization of Medical Knowledge:** Patients increasingly possess clinical sophistication rivaling non-specialist providers, enabled by online communities and direct-to-consumer diagnostics
- 2. Peer-to-Peer Triage as Healthcare Infrastructure:** Informal online networks are filling gaps in healthcare access, particularly for time-sensitive questions ("Should I go to ER tonight?") that formal systems handle poorly
- 3. Emotional Labor Externalization:** Healthcare systems constrained by time and reimbursement limitations increasingly rely on unpaid peer support to provide emotional care that physicians cannot
- 4. Platform Power and Responsibility:** Technology companies hosting health communities wield significant influence over health outcomes, raising questions about regulatory oversight and liability
- 5. The Quantified Patient:** Consumer health devices coupled with peer interpretation networks create parallel diagnostic infrastructures outside traditional medical gatekeeping

These trends suggest we are witnessing not merely the digitization of existing healthcare delivery models, but the **emergence of entirely new healthcare infrastructures** built on peer knowledge exchange, distributed triage, and algorithmic mediation. Understanding these systems, including their benefits, risks, and optimal governance, will be critical to 21st-century health policy.

6 Conclusion

This longitudinal analysis of 2,022 posts and 19,887 comments spanning 14 years (2012–2025) provides compelling evidence that the online pre-eclampsia community has undergone a fundamental transformation from a scattered collection of social support threads into a centralized, high-volume peer-triage ecosystem.

6.1 Core Contributions

Our research makes four primary contributions to digital health scholarship:

- 1. Temporal Dynamics:** We challenge the prevailing "COVID catalyst" hypothesis, demonstrating that the true inflection point for pregnancy-related online communities occurred in 2023–2025, not during the pandemic itself. This 2–3 year latency suggests that digital health adoption follows complex diffusion patterns mediated by network effects, platform maturation, and generational shifts.
- 2. Spatial Consolidation:** We document the "Great Migration" whereby 54.7% of discourse has concentrated into the dedicated `r/preeclampnia` subreddit, driven by lexical specialization (clinical monitoring vocabulary) and expertise agglomeration. This centralization demonstrates that high-stakes medical conditions consolidate into singular dominant forums despite platform fragmentation pressures.
- 3. Affective Infrastructure:** We quantify the "Buffer Effect"—a systematic $+0.328$ sentiment inversion between posts (distress signals) and comments (support responses), establishing that peer communities function as active emotional regulation systems, not

merely passive information repositories. The 19.4 percentage point shift from negative posts to positive comments represents robust community buffering with large effect size (Cohen's $h = 0.78$).

4. Functional Evolution: Through topic modeling, we reveal that the community's primary function has shifted from narrative sharing to quantitative peer-triage, with clinical monitoring discourse (Topic 3: "bp", "pressure", "readings") comprising 28% of all content. This "Clinical Fixation" demonstrates that patients are acting as distributed diagnostic networks, interpreting home-generated biometric data through crowdsourced expertise.

6.2 Practical Implications

For healthcare providers, our findings suggest that **patients are arriving at appointments with sophisticated clinical knowledge and community-validated concerns**. Rather than dismissing online communities as sources of misinformation, providers should engage with them as complementary support systems, offering guidance on interpreting peer advice and validating legitimate community concerns.

For platform architects and community moderators, the "sentiment cost of scale" (positivity declining from 57.8% to 46.8% as communities expand) indicates that **exponential growth requires active governance to maintain supportive culture**. Strategies including cultivation of success story pipelines, tiered expertise flair systems, and automated urgent-scenario detection can help preserve community buffering capacity despite volume increases.

For researchers, the pre-eclampsia community represents a valuable natural laboratory for studying **emergent healthcare infrastructure**, including peer triage systems, distributed diagnostic interpretation, and weak-tie emotional support, that increasingly supplement formal medical care. Understanding these systems' accuracy, safety, and health equity implications will be critical to 21st-century health policy.

6.3 The Road Ahead

As consumer health devices proliferate (home BP monitors, continuous glucose monitors, wearable sensors) and direct-to-consumer diagnostics expand (at-home genetic testing, telemedicine-enabled lab ordering), patients will generate ever-increasing volumes of clinical data outside formal healthcare settings. Online communities like **r/preeclampsia** serve as interpretive layers for these data, helping patients contextualize measurements, identify concerning patterns, and make care-seeking decisions.

This represents a fundamental shift from the 20th-century model of **physician-as-gatekeeper** to an emerging 21st-century model of **patient-as-co-diagnostician**, with online communities serving as consultation networks. Whether this shift improves or compromises care quality, equity, and safety will depend on how healthcare systems, technology platforms, policymakers, and patient communities navigate the complex integration of peer knowledge with professional expertise.

Our findings suggest cautious optimism: the pre-eclampsia community demonstrates robust buffering capacity, lexical sophistication, and functional specialization. However,

the sentiment cost of scale, potential for misinformation propagation, and systematic underrepresentation of high-risk populations (Black and Latinx women) demand ongoing vigilance and proactive governance.

The digital health revolution is not coming; it has arrived. The question is no longer whether patients will turn to online communities for medical advice, but how we can harness these communities' strengths while mitigating their risks. This research provides an empirical foundation for answering that question.

6.4 Final Reflection

Behind every data point in our corpus—every post, comment, sentiment score—lies a human story: a pregnant person experiencing terrifying symptoms, a survivor offering hard-won wisdom, a partner seeking to understand their loved one's crisis. Our quantitative methods risk abstracting away this humanity. Yet it is precisely by understanding these stories at scale that we can advocate for systems that better serve them.

The pre-eclampsia community is not perfect. It reflects the health inequities, platform biases, and medical uncertainty of the broader world. But it also represents something profound: strangers offering strangers hope at their darkest moments, knowledge translating into empowerment, and isolation transformed into connection. As healthcare continues its digital transformation, may we build systems worthy of the communities that patients have already created for themselves.