

母親向け SNS におけるコミュニティ特有表現の抽出と採用タイミングの予備的分析

内田ゆず¹ 高丸圭一² 乙武北斗³ 木村泰知⁴

¹ 北海学園大学 ² 宇都宮共和大学 ³ 福岡大学 ⁴ 小樽商科大学

yuzu@hgu.jp takamaru@kyowa-u.ac.jp

ototake@fukuoka-u.ac.jp kimura@res.otaru-uc.ac.jp

概要

母親向け SNS の Q&A データと BCCWJ を比較して母親コミュニティに特有な表現を抽出し、第一子の出産日（子どもの誕生日）を基点とした採用タイミングを分析する。対数オッズ比によりコミュニティに特有の表現 200 語を選び、ユーザごとの初出日と出産日との差の分布を求めた。その結果、離乳・保育園・イヤイヤ期など特定のライフイベントに対応して出現が集中する単峰型表現と、不妊治療や排卵日など出産前と出産後数年の二つのフェーズで出現する二峰型表現が確認された。

1 はじめに

近年、妊娠・出産・育児に関する情報共有の場として、母親向け SNS や Q&A コミュニティが広く利用されている [1][2][3]。これらの場では、一般的な書き言葉コーパスにはほとんど現れない、「10 倍粥」「イヤイヤ期」「赤ちゃん返り」「2 人目妊活」などの育児コミュニティ特有の言語表現が日常的に用いられている。こうしたコミュニティ特有表現は、単に専門用語というだけでなく、当事者が経験する具体的なライフイベントや育児の悩みと強く結びついていると考えられる。

産前産後の母親の課題を扱う研究では、子どもの月齢・年齢や妊娠週数に着目して、時期別の悩みや関心を調査するものが多い [4]。また、検索履歴を対象に、出産日付きの検索語ごとの確率分布を推定して産前産後の関心事を時期別に分析し、各月に推薦すべきキーワードを決定する「推薦カレンダー」を構成した研究も報告されている [5]。

本研究は、先行研究と問題意識を共有しつつ、対象とするデータと言語単位を大きく異なるものとする。すなわち、検索語ではなく母親向け SNS の

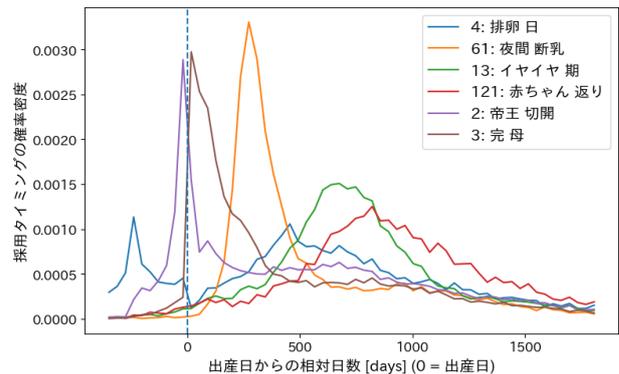


図1 代表的な特有表現における初出タイミング (例示)

Q&A テキストを対象とし、BCCWJ[6] との比較から母親コミュニティに特有な表現を自動抽出する。さらに、各ユーザがそれらの表現を初めて用いたタイミングを出産日からの相対日数として算出し、表現ごとの採用タイミング分布を推定・比較することで、母親向け SNS における言語使用の時期的な偏りを分析する。

図1に、代表的な特有表現の初出タイミング分布を例示する。同じ育児関連の話題であっても、初めて言及されやすい時期や分布の広がりや表現によって異なり、さまざまなパターンが見られる。本稿では、この差を出産日基点の初出分布として定量化するために、母親向け SNS から特有表現を抽出し、各表現の採用タイミング分布を推定・比較する枠組みを示す。

2 データと手法

2.1 データ

本研究で用いるデータは、コネヒト株式会社¹⁾が運営する女性向け Q&A アプリ「ママリ」の Q&A 投

1) <https://connehito.com/>

稿データである²⁾。2019年度～2020年度に投稿された質問約400万件、回答約2,400万件を利用し、質問文と回答文をまとめて「SNSテキスト」と呼ぶ。また、本研究では各表現の初出時刻を出産日を基点に扱うため、SNSテキストの投稿者のうち子ども情報が紐付いており出産日が取得できるユーザーに限定して分析を行う。該当ユーザーは約145万人である。

一般コーパスとしては、現代日本語書き言葉均衡コーパス(BCCWJ)を用いる。提供されている形態素情報付きテキストから、空白や句読点などの記号のみから成るトークンや、「笑」「w」などノイズとみなせるトークンを除去し、内容語(名詞・動詞・形容詞)のみのトークン列を得る。さらに、補助動詞や形式名詞は特有表現としての情報量が低いと判断し、集計対象から除外した。SNSテキストについても形態素解析器により同様のトークン列を作成した。

2.2 特有表現の抽出方法

SNSテキストとBCCWJのそれぞれについて、トークン列からbi-gramを抽出し、頻度を集計する。SNS側のbi-gram出現回数を c_{sns} 、全bi-gram出現回数を N_{sns} 、BCCWJ側をそれぞれ c_{bccwj} 、 N_{bccwj} 、語彙サイズ(異なりbi-gram数)を V とすると、加法スムージング($\alpha=1$)付きの確率推定は

$$p_{sns} = \frac{c_{sns} + \alpha}{N_{sns} + \alpha V} \quad (1)$$

$$p_{bccwj} = \frac{c_{bccwj} + \alpha}{N_{bccwj} + \alpha V} \quad (2)$$

で与えられる。SNS固有度の指標として、これらの比の対数

$$\log_odds = \log \frac{p_{sns}}{p_{bccwj}} \quad (3)$$

を用いる。SNS側の出現頻度が50以上であり、かつ \log_odds が大きいbi-gramを「母親向けSNSコミュニティに特有な表現(以下、特有表現)」とみなし、 \log_odds に基づいてソートした上位200語を本稿の分析対象とする。同じ手順で上位500語まで拡張した辞書も作成し、結果の解釈を補足する際に参照する。

2.3 採用タイミングの推定

各ユーザーについて、SNSテキスト内で特有表現を初めて使用した投稿のタイムスタンプを「初

出時刻」と定義する。質問と回答の両方を対象とし、ユーザー u 、特有表現 e に対して初出時刻を $first_date(u, e)$ とする。本稿では単純化のため、ユーザーに紐づく子どもレコードのうち、第一子の出産日のみを用い、その日付を $birth_date(u)$ とする。

出産日を時間軸の基点とし、 $first_date(u, e) - birth_date(u)$ を特有表現の採用タイミングと定義する(単位は日)。極端な外れ値の影響を抑えるため、相対日数が $[-365, 3650]$ の範囲に入る表現のみを分析対象とした。各特有表現について、採用タイミングの分布から中央値、四分位範囲、出産前に初出する割合、出産から3年以上経過してから初出する割合などの統計量を算出し、ヒストグラムや確率密度の可視化により時間分布の形状を観察する。なお、これらの統計量の具体例を、付録(表1,表2)に示す。

3 結果

3.1 抽出された特有表現の概要

2章の手順により、母親向けSNSとBCCWJのbi-gram頻度から \log_odds を算出し、SNS側で出現頻度が高く、かつ \log_odds が大きい順に並べた上位200語を特有表現として得た。抽出された表現には、「生後ヶ月」「帝王切開」「完母」「ギャン泣き」「排卵日」「イヤイヤ期」「倍粥」「慣らす保育」「赤ちゃん返り」「目妊活」「産休育休」など、妊娠・出産医療、授乳・離乳、睡眠、保育園入園、幼児期の行動、第二子妊活、雇用制度といった出産・育児イベントに密接に関連する語が多く含まれていた。以下では、これらのうち代表的な表現について、採用タイミング分布の具体例を示す。

3.2 採用タイミング分布のパターン

3.2.1 単峰型の分布を示す表現の例

まず、採用タイミング分布が単一の顕著なピークを持つ単峰型の例として、「倍粥」および「イヤイヤ期」を示す。

図2は、「倍粥」の採用タイミングのヒストグラムである。横軸は第一子の出産日からの経過日数、縦軸はその区間で初めて当該表現を用いたユーザー数である。分布は、生後半年以降に立ち上がり、生後半年～1年程度の範囲に山が集中する形状を示した。

図3は、「イヤイヤ期」のヒストグラムである。

2) 個人を識別する情報は一切含まれない

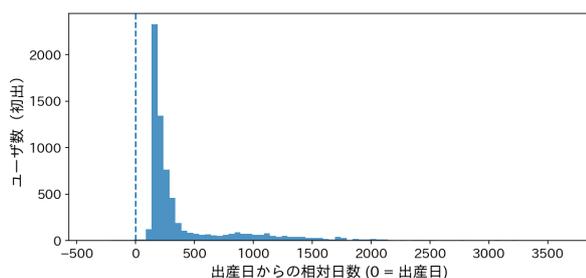


図2 「倍粥」における採用タイミング分布

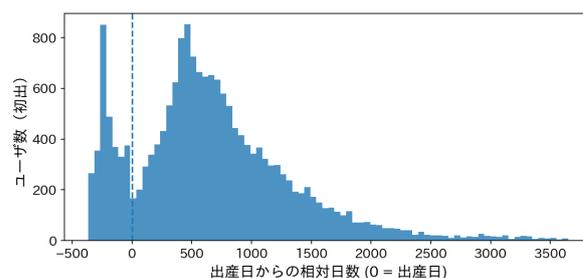


図4 「排卵日」における採用タイミング分布

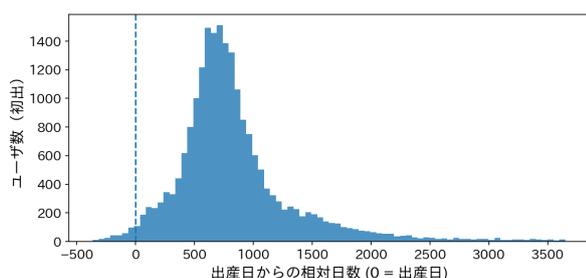


図3 「イヤイヤ期」における採用タイミング分布

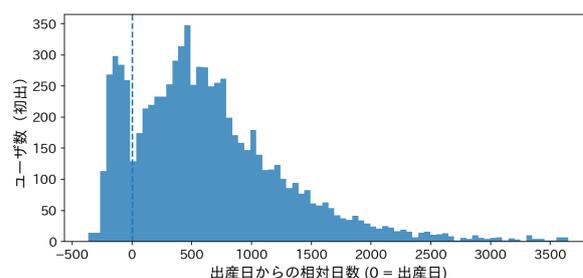


図5 「産休育休」における採用タイミング分布

生後0~400日では初出は少ないが、生後600~1,000日付近に大きな山が現れ、中央値は約740日(2歳前後)である。また、分布は4~5歳相当の領域まで長いテールを持つ。

3.2.2 二峰型の分布を示す表現の例

次に、採用タイミング分布が複数のピークを持つ表現の例として、「排卵日」と「産休育休」を示す。

図4は、「排卵日」のヒストグラムである。分布は出産前(負の領域)にピークを持つと同時に、出産後1~2年の範囲にも別の山が現れる。

図5は、「産休育休」の分布である。出産日前後にかけて第1のピークがあり、出産から1~3年後にも第2のピークが現れる。

3.3 上位500語への拡張結果

本稿では分析対象を上位200語の特有表現辞書としたが、同じ手順でlog_oddsに基づく上位500語まで拡張した辞書も構築し、採用タイミングの基本的な統計量を算出した。その結果、200語で観察された分布形状の傾向(単峰型・二峰型の混在や、フェーズに沿った配置)が、より広い語彙集合に対しても概ね維持されることが確認された。

具体的には、離乳フェーズに関しては、「倍粥」に加えて「離乳食つく」「離乳食後」「離乳食ミルク」など複数の表現が追加され、いずれも生後6~

10ヶ月付近に山が集中する分布を示した。妊娠期については、「初マタ」「妊娠週」「胎嚢確認」などの語が加わり、出産前に初出する割合が高かった。さらに、「上2人」「子2人」「ジュニアシート」など複数の子ども・園生活に関連する語は、出産から3年以降に初出する割合が高かった。

4 考察

4.1 単峰型表現の解釈

3章で示した単峰型の例は、子どもの発達段階に対応して、出産日からの一定の時期に初出が集中しやすいことを示している。ここでは、「倍粥」「イヤイヤ期」について、分布形状の背景を考察する。

「倍粥」 一般的に、離乳食は5~6ヵ月頃に10倍粥から開始し、7~8ヵ月頃に7倍粥へ、9~11ヵ月頃には5倍粥から軟飯へと、月齢に応じて濃度や回数を段階的に上げていくことが推奨されている。本稿での「倍粥」は、「10倍粥」「7倍粥」「5倍粥」などを含んでいると考えられる。図2で見られた「生後半年以降の急な立ち上がり」と「生後半年~1年程度への集中」は、離乳食の開始そのものというよりも、食事回数や粥の濃度を移行させる時期にかけて、母親が「何倍粥に進めてよいか」「どのタイミングで軟飯に移行するか」といった具体的な悩みを言語化し始めることを反映している可能性がある。

すなわち、「倍粥」という表現は、「離乳食の進め方・ステップアップ」に関する情報ニーズが高まる時期を示す指標として機能していると解釈できる。

「イヤイヤ期」 歩行開始期にあたる1歳後半から2歳前半にかけての時期は、一般的に子どもの反抗や自己主張が強くなり、「第一次反抗期」あるいは「イヤイヤ期」と呼ばれる[7]。また、0歳～5歳のイヤイヤ期の子どもを持つ女性を対象としたオンライン調査³⁾では、5歳になってもおよそ4人に1人が「イヤイヤ期」と認識しているという結果が報告されている。図3では、中央値が約2歳前後である一方、4～5歳相当まで長いテールが観察された。これは、上記調査が示す「イヤイヤ期」の認識の広がりや統合的である。このことから、採用タイミング分布は、発達心理学的に定義された期間そのものというよりも、「親にとってのイヤイヤ期」の主観的な時間幅が、言語表現の初出として可視化されたものと解釈できる。

4.2 二峰型表現の解釈

二峰型の例として示した「排卵日」「産休 育休」は、出産日前後と出産後数年という複数のフェーズで初出の山を持つ。

「排卵日」 出産前（負の領域）のピークは、第一子妊娠前後の排卵日計算や排卵検査に関する言及を反映している可能性がある。また、出産後1～2年の範囲に現れる第2の山は、第二子妊活の開始に際して再び当該表現が用いられる状況を反映している可能性がある。

「産休 育休」 出産日前後にかけての第1ピークは、第一子出産に伴う産前産後休業・育児休業の取得に関する情報収集・相談だと考えられる。一方、出産後1～3年後の第2ピークは、第二子以降の妊娠・出産に際して、再び「産休 育休」に関する情報収集や意思決定が行われる状況を反映している可能性がある。

以上より、同じ表現が時間を隔てて再び出現することは、「妊娠・出産」や「制度利用」が1回のイベントではなく、計画・治療・制度調整を含むプロセスとして複数回経験されうることを示唆する。この観点から、二峰型表現は、単一フェーズの指標というよりも、複数のフェーズにまたがる反復的な意思決定・調整を伴うテーマの指標として捉えられる。

3) 博報堂こそだて家族研究所「イヤイヤ実態調査」
<https://www.hakuhodo.co.jp/news/info/45186/>

4.3 辞書拡張の有効性と応用可能性

3.3の結果から、上位500語へ拡張しても、特有表現の初出は妊娠期（出産前）、乳児期～幼児期（出産後早期）、就園期・第二子以降（出産後数年）といったフェーズに沿って偏る傾向が概ね維持された。すなわち、語彙集合を広げても「どの表現がどの時期に立ち上がりやすいか」という対応関係は大きく崩れず、特有表現は頻度だけでなく初出時期によっても特徴づけられる。

本研究は、投稿を「離乳食」「睡眠」等のカテゴリに分けて月齢別件数を集計するのではなく、特有表現ごとに「その表現がユーザに初めて現れる時期」を集計する点に特徴がある。分布の立ち上がりやテールの長さは、発達段階の区分そのものというよりも、母親が当該テーマについて迷いや負担を強く意識する主観的な時間幅を反映している可能性がある。このように、特有表現の時間特性を通じて「親にとっての育児イベント」の立ち上がりや持続を間接的に観測することで、行政や医療・保健機関による月齢・年齢に応じた情報提供や相談・レコメンドのタイミング設計を検討する際の補助情報となりうる。

5 まとめと今後の課題

本研究では、母親向けSNSとBCCWJのbi-gram頻度を比較して母親コミュニティにおける特有表現の辞書を構築し、出産日を基点とした採用タイミング分布を分析した。「倍粥」「イヤイヤ期」などは離乳・幼児期といった特定フェーズで集中的に現れる一方、「排卵日」「産休 育休」などは第一子妊娠前後と第二子以降の妊娠・出産期の双方で再び立ち上がる二峰型の分布を示し、上位500語への拡張においても同様の時間構造が確認された。

一方で、本研究はbi-gramに対象を限定し、第一子の出産日のみを時間軸の基点とするなど、データの扱いを単純化している。また、比較対象にBCCWJを用いることで「一般的な書き言葉」との対比は可能になっているものの、ウェブ上の口語や他のSNSにおける使用頻度までは十分に反映できていない。

今後は、より長いフレーズへの拡張、他コーパスとの比較、投稿頻度や属性情報との統合を通じて、時間軸付き語彙としての整理を精緻化し、育児支援や情報提供の実践に結び付けていくことが課題である。

謝辞

本研究の遂行にあたり、貴重なデータを提供していただいたコネヒト株式会社様に感謝の意を表す。本研究は JSPS 科研費 24K15196 の助成を受けたものである。

参考文献

- [1] 中島千英子, 永井由美子. 母親の育児情報源としての sns 利用に関する調査. 大阪教育大学紀要. 人文社会科学・自然科学, Vol. 68, pp. 41–49, 02 2020.
- [2] 平松紀代子. 育児情報源と育児情報の獲得状況に関する一研究 — 0 歳児を持つ母親を対象に—. 家族関係学, Vol. 43, No. 0, pp. 65–77, 2024.
- [3] 真鍋健, 明田楓可, 森菜津子. 発達支援に関わる情報に保護者はどのようにアクセスしているのか: 自治体からの情報提供と sns を含めたソーシャルメディアに注目したアンケート調査. 千葉大学教育学部研究紀要, Vol. 73, pp. 381–388, 03 2025.
- [4] 申沙羅, 山田和子, 森岡郁晴. 生後 2～3 か月児がいる母親の育児困難感とその関連要因. 日本看護研究学会雑誌, Vol. 38, No. 5, pp. 33–40, 2015.
- [5] 岩永二郎, 吉田晏大, 高橋彩音. 検索語の確率分布に基づく時期別の母親の関心事に合わせた情報推薦—特集 データ解析コンペティション: コミュニケーションサイトの分析. オペレーションズ・リサーチ: 経営の科学, Vol. 68, No. 2, pp. 47–55, 2 2023.
- [6] 前川喜久雄 (監修) 山崎誠 (編). 講座日本語コーパス 2 書き言葉コーパス-設計と構築-. 朝倉書店, 2014.
- [7] 坂上裕子. 歩行開始期幼児の反抗・自己主張への母親の対応をめぐる経年変化: 2000 年と 2020 年の質問紙調査の結果の比較. 教育人間科学部紀要, Vol. 13, pp. 97–115, 03 2022.

付録：コミュニティ特有表現の採用タイミングの概要

3章・4章で分析対象とした表現のうち、採用タイミングが比較的狭い範囲に集中するもの（表1）、出産前後など複数のフェーズで出現するもの（表2）の代表例を示す。各表現について、 n_{users} は当該表現を一度以上用いたユーザー数、median は初出相対日数の中央値、IQR は四分位範囲、 p_{neg} は出産前に初出する割合、 p_{post3y} は出産から3年以上経過してから初出する割合、 n_{peaks} は平滑化したヒストグラムにおけるピーク数、 \log_odds は母親向け SNS と BCCWJ の bi-gram 出現確率比の対数である。

表1 鋭いピークをもつ表現（20語）

| ID | bi-gram | n_{users} | median | IQR | p_{neg} | p_{post3y} | n_{peaks} | \log_odds |
|-----|---------|-------------|--------|--------|-----------|--------------|-------------|--------------|
| 174 | 倍粥 | 6920 | 219 | 192.00 | 0.00 | 0.10 | 1 | 10.90 |
| 140 | 初マタ | 9173 | -49 | 315.00 | 0.70 | 0.07 | 1 | 11.06 |
| 58 | 一半 | 11168 | 581 | 380.00 | 0.01 | 0.19 | 1 | 11.57 |
| 146 | 半検診 | 8398 | 614 | 382.00 | 0.01 | 0.20 | 1 | 11.01 |
| 13 | イヤイヤ期 | 21073 | 739 | 425.00 | 0.02 | 0.19 | 1 | 12.29 |
| 34 | ベビーフード | 14033 | 351 | 460.00 | 0.01 | 0.14 | 1 | 11.89 |
| 61 | 夜間断乳 | 8972 | 372 | 498.00 | 0.00 | 0.14 | 1 | 11.55 |
| 173 | 育休延長 | 6021 | 425 | 518.00 | 0.02 | 0.15 | 1 | 10.91 |
| 156 | 目妊活 | 7110 | 653 | 525.00 | 0.04 | 0.18 | 3 | 10.97 |
| 164 | 2人妊娠中 | 9897 | 709 | 545.00 | 0.04 | 0.22 | 1 | 10.93 |
| 152 | 突発発疹 | 8380 | 550 | 545.25 | 0.01 | 0.19 | 1 | 10.99 |
| 199 | パンツタイプ | 8012 | 274 | 565.25 | 0.02 | 0.14 | 1 | 10.86 |
| 171 | 新生児頃 | 9330 | 255 | 571.00 | 0.02 | 0.12 | 1 | 10.91 |
| 121 | 赤ちゃん返り | 8563 | 917 | 578.00 | 0.02 | 0.35 | 1 | 11.14 |
| 95 | 抱っこ寝る | 13215 | 268 | 581.00 | 0.00 | 0.12 | 1 | 11.30 |
| 29 | ヶ月息子 | 21768 | 337 | 585.00 | 0.00 | 0.12 | 4 | 12.04 |
| 19 | ヶ月娘 | 23767 | 327 | 587.50 | 0.00 | 0.12 | 2 | 12.21 |
| 78 | 月保育園 | 14634 | 494 | 588.00 | 0.02 | 0.17 | 1 | 11.41 |
| 157 | 回起きる | 9896 | 311 | 592.00 | 0.03 | 0.14 | 1 | 10.97 |
| 131 | 2人妊娠 | 11207 | 686 | 592.00 | 0.04 | 0.22 | 2 | 11.11 |

表2 二峰性／多峰性をもつ表現（20語）

| ID | bi-gram | n_{users} | median | IQR | p_{neg} | p_{post3y} | n_{peaks} | \log_odds |
|-----|---------|-------------|--------|---------|-----------|--------------|-------------|--------------|
| 4 | 排卵日 | 17938 | 582 | 763.00 | 0.17 | 0.22 | 4 | 12.87 |
| 10 | 排卵検査 | 9279 | 609 | 692.50 | 0.17 | 0.21 | 3 | 12.34 |
| 11 | 基礎体温 | 11509 | 553 | 811.00 | 0.21 | 0.20 | 3 | 12.32 |
| 14 | 不妊治療 | 12302 | 480 | 939.00 | 0.25 | 0.19 | 5 | 12.28 |
| 18 | 妊娠前 | 25147 | 288 | 806.00 | 0.29 | 0.15 | 4 | 12.25 |
| 23 | 心拍確認 | 16376 | 481 | 1011.00 | 0.30 | 0.18 | 2 | 12.12 |
| 43 | 体外受精 | 4795 | 398 | 970.50 | 0.31 | 0.17 | 4 | 11.73 |
| 50 | 妊娠初期 | 16764 | 440 | 930.00 | 0.31 | 0.17 | 3 | 11.63 |
| 56 | 自宅安静 | 9760 | 507 | 977.00 | 0.31 | 0.19 | 2 | 11.59 |
| 84 | 化学流産 | 5127 | 603 | 905.50 | 0.23 | 0.22 | 4 | 11.39 |
| 108 | 予定日前 | 9863 | 559 | 891.00 | 0.21 | 0.21 | 3 | 11.23 |
| 113 | 陽性反応 | 9018 | 554 | 885.00 | 0.24 | 0.20 | 3 | 11.21 |
| 160 | 里帰り出産 | 9462 | 343 | 817.00 | 0.31 | 0.13 | 4 | 10.96 |
| 168 | 産休育休 | 7799 | 537 | 785.00 | 0.17 | 0.20 | 5 | 10.92 |
| 178 | 週頃 | 8867 | 362 | 930.00 | 0.37 | 0.17 | 2 | 10.90 |
| 185 | 自然妊娠 | 5278 | 541 | 895.00 | 0.22 | 0.21 | 6 | 10.88 |
| 191 | 妊婦健診 | 8732 | 461 | 927.00 | 0.32 | 0.18 | 3 | 10.88 |
| 200 | 今妊娠 | 10574 | 458 | 978.00 | 0.36 | 0.18 | 2 | 10.86 |
| 281 | 妊婦検診 | 7895 | 478 | 979.00 | 0.33 | 0.19 | 3 | 10.68 |
| 482 | エコー見る | 6489 | 447 | 1013.00 | 0.34 | 0.20 | 2 | 10.33 |