



The real-world applications of the symptom tracking functionality available to menstrual health tracking apps

Tatheer Adnan^a, Brent A. Coull^{a,b}, Anne Marie Jukic^c
and Shruthi Mahalingaiah^{a,d}

Purpose of review

The goal of this review was to evaluate whether the fields available in iOS mobile phone apps for recording menstrual cycle symptoms are able to be harmonized across platforms for potential use in research, such as aggregated data analysis.

Recent findings

Symptom tracking capabilities are a common feature among menstrual health apps but have been the subject of limited investigations. Apps differ with respect to which symptoms are included and the rationale for these differences is unclear. Epidemiologic studies are poised to incorporate these data; however, a thorough exploration of symptom tracking functionality across apps is required.

Summary

Our review finds that the language used to describe symptoms and the specificity with which symptoms are collected varies greatly across the most used iOS tracking apps. Although some apps allow numerical and qualitative description of symptoms, such as sleep and mood, most simply record the presence or absence of a symptom. Collaborative efforts between clinicians and researchers to guide what and how data is collected may allow tracking apps to realize their potential diagnostic applicability. Regardless, with the increasing use of menstrual health tracking apps, it is imperative that data retrieved from such apps can realize its potential in the research and clinical ecosystems.

Keywords

cycle-tracking apps, digital diagnostics, menstrual health symptoms

INTRODUCTION

The global women's health app market was valued at more than two billion in 2020. The segment of menstrual health apps dominates the field contributing to 40% of the revenue share, and is estimated to have a strong, growing presence from 2021 to 2028 [1].

Studies suggest that the primary motivators for the use of self-tracking menstrual health apps include identification of bleeding irregularities, preparing for upcoming periods, improving knowledge of menstrual cycles, facilitating conversations with healthcare professionals, verification of menstrual experiences, and timing intercourse to either avoid or facilitate conception [2[•]]. Given these motivators, most menstrual health apps integrate personal health information, calendar tracking, and menstrual symptom collection.

The existing literature pertaining to menstrual tracking apps can be categorized broadly into those that develop and employ tools to holistically

evaluate, or rate, menstrual health tracking apps [3[•],4], create large app-specific databases of menstrual cycles to examine associations between menstrual cycle characteristics [5[•],6^{••}], and conduct

^aDepartment of Environmental Health, ^bDepartment of Biostatistics, Harvard T.H. Chan School of Public Health, Boston, Massachusetts, ^cEpidemiology Branch, Division of Intramural Research, National Institute of Environmental Health Sciences, National Institutes of Health, Research Triangle Park, North Carolina and ^dDepartment of Obstetrics, Gynecology and Reproductive Biology, Massachusetts General Hospital and Harvard Medical School, Boston, Massachusetts, USA

Correspondence to Tatheer Adnan, Visiting Scholar, Harvard School of Public Health. 677 Huntington Avenue, Boston, MA 02115, USA. Tel: +1 347 977 9492; e-mail: tatheeradnan@gmail.com

Curr Opin Endocrinol Diabetes Obes 2021, 28:574–586

DOI:10.1097/MED.0000000000000682

This is an open access article distributed under the terms of the Creative Commons Attribution-Non Commercial-No Derivatives License 4.0 (CCBY-NC-ND), where it is permissible to download and share the work provided it is properly cited. The work cannot be changed in any way or used commercially without permission from the journal.

KEY POINTS

- Apps predominantly collect physical symptoms, using scales (light, medium, and heavy; mild, moderate, severe) to characterize the experience. There is, however, significant discrepancy amongst the terminology and specificity with which different symptoms are collected.
- Data integration capabilities are currently underdeveloped. Symptom-related data cannot be shared between independent health tracking apps and Apple Health.
- The low granularity with which apps are currently collecting symptoms does not allow apps to realize their diagnostic potential. Pain, for instance, is only captured as present or absent in menstrual health tracking apps or ranked along a vague mild–severe scale.
- Users are not consistent and thorough in their symptom tracking practices. The variation in user behavior means that symptom tracking data is riddled with missing and incomplete data.
- Researchers and clinicians must collaborate to render symptom field options relevant and harmonizable across platforms for future aggregated data analysis.

qualitative analyses of user motivations and behaviors [2[¶]]. Data analyzed from app-specific repositories has already challenged the textbook definition of normal menstrual cycle characteristics, including

the notion of a 28-day cycle [7[¶]]. To our knowledge there, however, is no comprehensive review of the categories of symptoms included across menstrual health apps and the extent to which the symptoms included are clinically informed and validated. Although there have been recent efforts to standardize the language used to characterize menstrual experiences, the degree to which this has been adopted by menstrual health apps has not been explored [8]. The goal of this review, therefore, was to systematically catalog symptom field options by menstrual tracking apps to understand whether these symptoms and the language used to describe them are similar across apps, ultimately using the data collected to evaluate whether symptom field options are harmonizable across platforms for future aggregated data analysis.

METHODS

Mobile apps reviewed were identified by searching the Apple iTunes Store. Driven by the invent of the Apple Cycle Tracker, a software compatible with Apple's Health App on iOS 13 and Watch OS 6 update, in June 2019, the current review was limited to the iOS platform. The following search terms were used: period tracking/monitoring, menstrual tracking/monitoring, and cycle tracking/monitoring to identify relevant apps. The use of these search terms saturated results, such that no new apps were revealed towards the end of the process. Of the apps remaining, those with greater than 1 000 000 downloads,

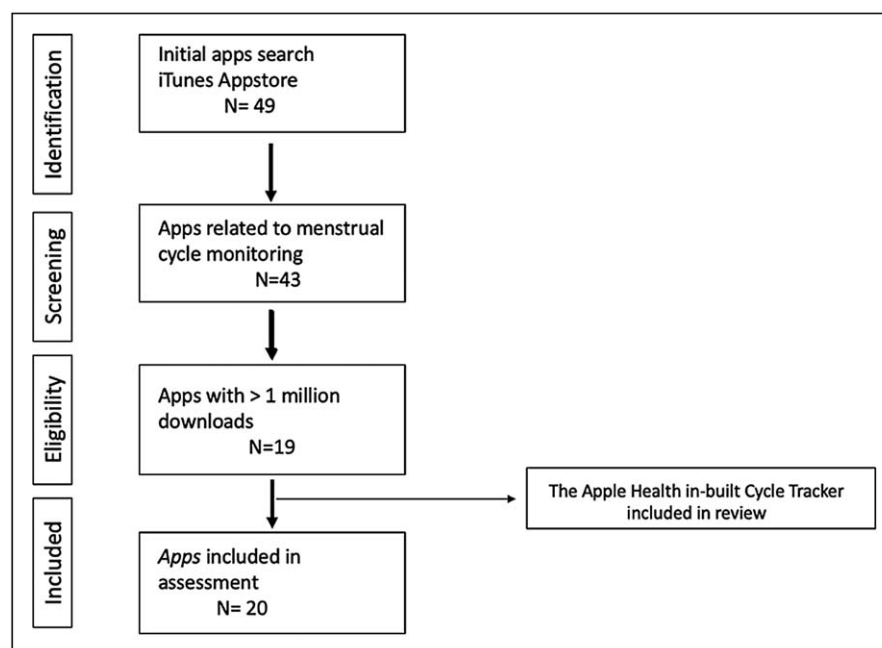


FIGURE 1. The App Selection Process. The PRISMA diagram for the systematic review of the iTunes Appstore detailing the search terms, screening process, and inclusion criteria.

19 apps, were included in the analysis (Fig. 1). In addition, the in-built Apple Health Cycle Tracker was incorporated into the study ($n = 20$). Although the review was conducted on an iOS platform, 15 of the 20 apps reviewed are also available on an Android operating system. A list of the apps included (see Table, Supplemental Digital Content 1, <http://links.lww.com/COE/A30> provides names, acronyms, and icons for the apps reviewed) and some characteristics of the apps (see Table, Supplemental Digital Content 2, <http://links.lww.com/COE/A30> presents information about app popularity and design) are presented in supplemental material, <http://links.lww.com/COE/A30>. Apps meeting the inclusion criteria were downloaded onto an iPhone, and the features of each piloted through a simulated cycle. Data collected for each app included: the number and category of menstrual cycle symptoms, defined as any observation the app allowed a user to record, the language used to record the metric, the app structure of metric reporting (multiple choice, free text entry) and the level of detail that can be recorded for each metric – for example, is the metric recorded as present or absent, or is it recorded in terms of severity or frequency, and the ability of apps to sync data across different health apps and wearables.

RESULTS

Number and category of menstrual symptoms recorded

All apps reviewed ($n = 20$) provided at least 10 symptoms users could track. The apps providing the most extensive list of metrics included Ovia and WomanLog, which offered users more than 100 symptoms to track. Two apps were customizable and allowed users to create and add symptom tags to the default options presented (Clue, F. F). Although not allowing the user to create symptom tags, many apps include a free journaling feature encouraging users to record information not captured readily by the default interface (My Calendar, P.T., P.C., Ovia, M.P.T, P.D, Flo, Maya, LadyTimer, PepApp, MyDaysX, F.F, WomanLog; $n = 13$).

Certain apps organized metrics into categories. Across the different apps, metrics are not always attributed to the same category. For instance, while certain apps categorize fatigue as an independent physical symptom, some present it as an option-describing mood. Therefore, a comprehensive categorization of metrics into larger themes of body metrics, period-related metrics, cervical state, physical symptoms, emotional symptoms, behavioral attributes, and activity levels is presented in Table 1. Physical symptoms constitute the largest category of symptoms.

Table 1. A comprehensive list of metrics collected across all menstrual health tracking apps

	Symptoms	Number of apps
Body metrics	Weight	14
	Basal BT	11
	Body temperature	8
Period	Spotting	17
	Bleeding; flow	14
	Discharge	9
	Collection method	2
Cervical state	Cervical fluid/mucus quality	12
	Cervical opening	8
	Cervical firmness/texture	7
	Cervical height/position	3
	Cervical mucus amount	2
Physical symptoms	Headache	20
	Cramps	19
	Skin/acne	19
	Backache/back pain	18
	Breast sensations	18
	Diarrhea	16
	Bloating	16
	Nausea	15
	Constipation	14
	Dizziness	11
	Migraine	10
	Hot flashes	10
	Body aches	8
	Muscle pain/joint pain	8
	Itchiness	8
	Ailment/sick	8
	Neckaches	7
	Chills	6
	Ovulation pain	6
	Bowel movement	6
	Hair loss	6
	Pelvic pain	5
	Stool	5
	Rashes	4
	Night sweats	4
	Frequent urination	4
Shoulder aches	3	
Abdominal cramps	2	
Vaginal pain	2	
Pain when urinating	2	
Facial hair	1	
Emotional health	Mood/moodiness	14
	Fatigue/tired	14
	Emotions	10
	Mental	8

Table 1 (Continued)

	Symptoms	Number of apps
	Motivations	6
	Stress	6
	Anxiety	6
Behavioral	Sleep patterns	15
	Cravings/appetite	13
	PMS	11
	Digestion/indigestion	10
	Pain during sex	6
	Sex drive	5
	Activities	Exercise
	Meditation	6
	Nutrition	2

The metrics have been organized into larger categories of symptoms. However, each metric is accompanied with the number of apps that make its reporting available to users independent of whether and which category the metric was placed within by a particular app. BT, basal body temperature; PMS, premenstrual syndrome.

Previous research indicates that even amongst apps that allow for the collection of symptoms across numerous categories, period flow, emotional state, and experience of pain are the most frequently tracked categories [6²²]. Effectively, all apps reviewed in this study collected at least one flow-related metric, physical/pain measure, and behavioral symptom,

whereas only half of the apps collected symptoms across all categories (Fig. 2). Nonetheless, on average, 48% of all metrics collected by an app are physical symptoms, intended to characterize the pain experienced by the user. Skin conditions, flow-related metrics, and symptoms that characterize body pain and gastrointestinal discomfort are among the physical features collected by the most apps (Fig. 3).

The language of symptoms

Many apps used scales (light, medium, and heavy) to track menstrual flow. There was, however, significant discrepancy amongst the terminology used to collect metrics related to bleeding days. Bleeding, flow, spotting, and discharge were used interchangeably. Certain apps (Clue, SpotOn, FF; $n = 3$) collected spotting as a qualifier of bleeding and flow (e.g. spotting, light, medium, and heavy) whereas most collected both 'spotting' and 'flow' (Glow, My Calendar, P.C, Eve, Ovia, Ladytimer; $n = 6$). 'Flow', in such cases, was reserved for bleeding days. The remaining collected either spotting or flow. Similarly, a variety of terms, such as breast tenderness, sensitivity, swelling, sore nipples, breast pain, and sore breasts were used to describe breast sensations. Although most apps chose a single term to capture the concept, certain apps collected a subset of breast sensations without providing explanations to

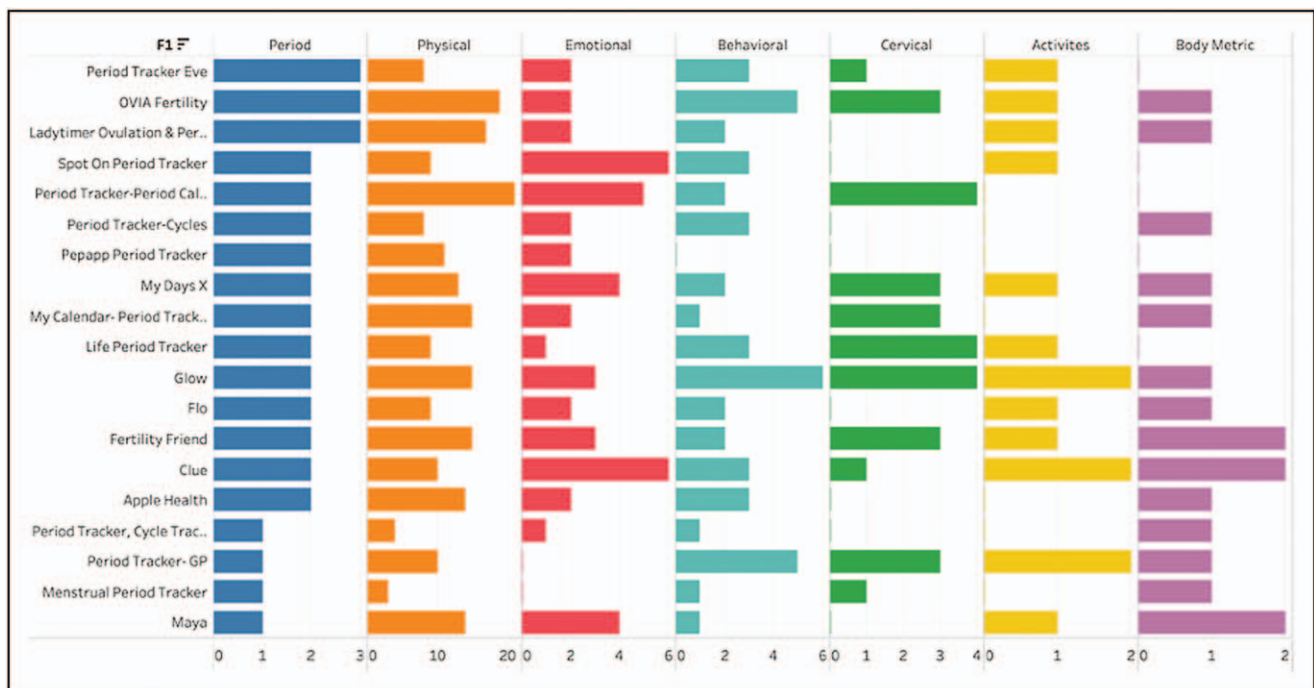


FIGURE 2. Number of metrics within each category of symptoms recorded by the apps reviewed. All the apps collect multiple metrics related to bleeding days and the physical pain experienced by the menstruator. Emotional health and behavioral changes, such as increased/decreased sex drive and appetite, are also commonly recorded.

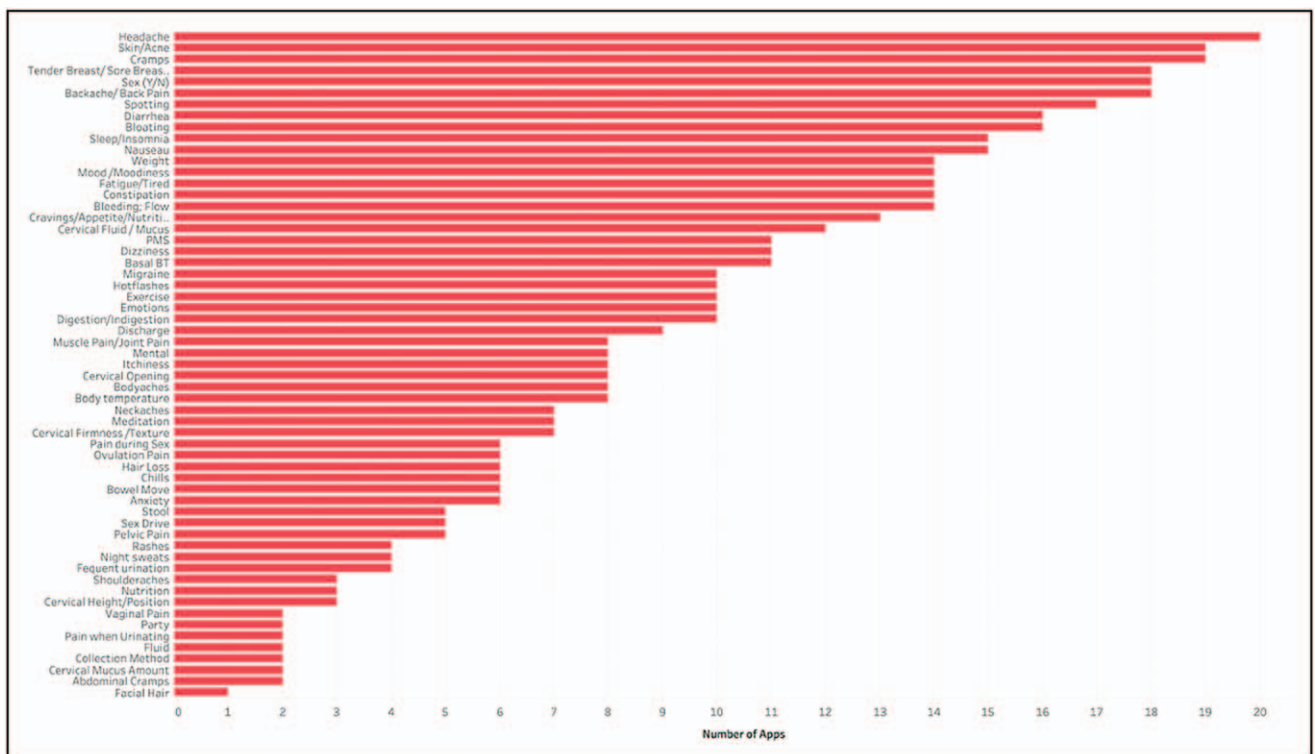


FIGURE 3. Distribution of symptom collection across reviewed apps. Apps predominantly collect physical symptoms. Body aches, breast sensitivity, and gastrointestinal states are amongst the most frequently collected physical symptoms.

elucidate the subtle difference among the symptoms presented (My Calendar, Period Calendar, Ovia, MyDays X, FF, WomanLog; $n = 6$). Both My Calendar and Period Calendar, for instance, presented breast sensitivity and tenderness as potential experiences. Although less ambiguous, a variety of terms were used to capture concepts, such as sex and sex drive, dietary habits, digestion, and period-related pain. The different terms that are used to capture a particular concept are presented in Table 2.

The language of symptoms pertaining to digestive state and cervical mucus quality was the most consistent. Most apps allowed for the recording of whether a user was experiencing ‘bloating, constipation, diarrhea, nausea, or gaseousness’. Certain apps included an additional metric for the presence or absence of indigestion, also referred to as dyspepsia. The language used to qualify cervical mucus quality, nonetheless, had the highest consistency across the different apps reviewed. All apps

Table 2. High variance in language used to capture metrics

Concept	Terms used across apps (N = 20)
Bleeding day measures	Period, bleeding, spotting, flow, discharge
Uterine pain/discomfort	Ovulation pain, pelvic pain, vaginal pain, cramps, abdominal cramps
Digestion	Indigestion, dyspepsia, stomach ache, upset stomach, heartburn
Breast sensations	Tender breasts, breast sensitivity, sore nipples, breast pain, swollen breasts, aching nipples, breast changes, breast engorgement
Dietary habits	Hungry, indulged, cravings (salty, sweet, etc.), increased/decreased appetite, ate junk
Sex	Unprotected, protected, with condom, without condom, banana free, withdrawal, had sex, multiple sessions, no sex, masturbation, kissing, make out session, pull out, intercourse, insemination
Sex drive	Do me now, I’m down, MIA, not today, high sex drive, strong sex drive

Across all 20 apps reviewed, a variety of terms are used to capture the same concept. For instance, ‘tenderness, sensitivity, swelling, sore nipples’ are all terms used to understand the breast sensations a user is experiencing. While indigestion, or dyspepsia, are commonly used to understand digestive issues, certain apps use more colloquial terms such as stomach ache and heartburn.

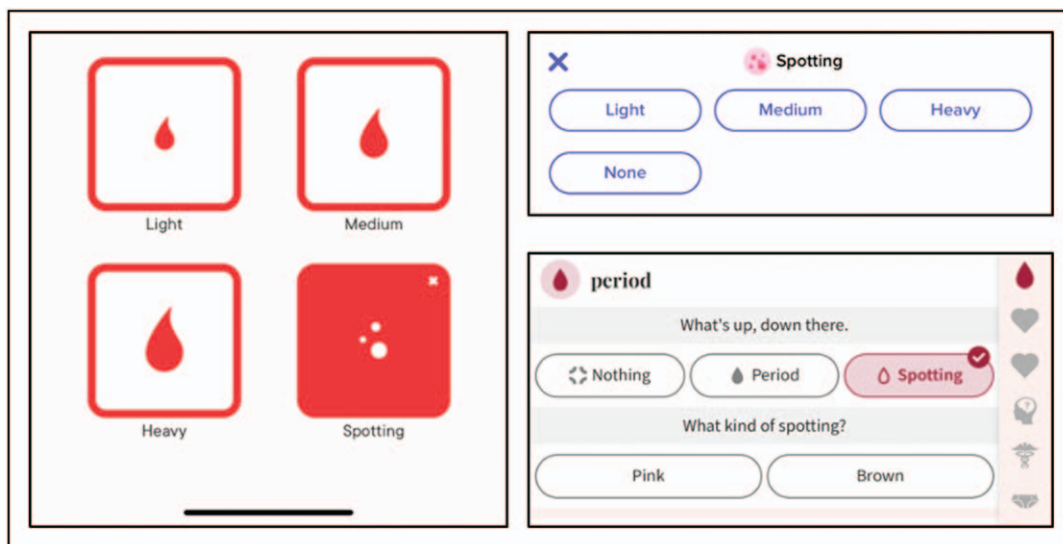


FIGURE 4. Levels of granularity for symptom collection. Metrics, such as spotting can be recorded at varying degrees of granularity. In the leftmost image, App A (Clue) records the presence or absence of spotting, App B (Glow) qualifies spotting along a none-heavy scale, and App C (Ovia) allows users to provide information about the color of the spotting.

presented ‘Dry, Sticky, Creamy, Watery, and Egg White’ as descriptions of mucus quality.

Despite the use of overlapping, ambiguous terminology, few apps provided definitions of the metrics being collected. Although not providing definitions, both Flo and Clue provided insights into the symptom once a user had indicated its presence. Flo, for instance, provided information about potential causes, symptom prevalence, and mitigating factors, and behaviors. Fertility Friend was the only app to both provide definitions of a metric and explain how it should be recorded.

Granularity of symptom recording

Apps differ in the amount of detail that is recorded for symptoms. Some metrics merely captured the presence or absence of a symptom, some were ranked by severity along a scale corresponding roughly to light, medium, and heavy, and some allowed additional quantitative/qualitative descriptions. An example of the hierarchical granularity with which a symptom, such as spotting, was recorded can be seen in Fig. 4. For spotting, the least granular apps recorded the presence/absence of spotting. Those that are more granular allowed the user to qualify spotting along a scale, and the most granular further incorporated descriptions, such as color and texture of the spotting. Most apps collected different symptoms at differing levels of specificity. Symptoms, such as ‘cramps’, ‘headaches’, ‘backache’, ‘constipation’, ‘diarrhea’, ‘bloating’, ‘acne’ were commonly collected at the present/

absent level. Mood was collected at the most granular level, with users of most apps having the ability to choose among 10+ moods. Metrics that allowed for additional qualitative/quantitative description included weight, activity levels, sleep cycle, and changes in appetite.

Certain apps collected a single metric in multiple ways. For instance, My Calendar allowed users to both input a numerical value for weight and indicate the presence/absence of weight gain. Similarly, Glow recorded quantitative values for sleep, while also allowing a user to qualify insomnia, if experienced, along a scale. The maximum granularity with which the metrics analyzed are recorded across the different apps reviewed can be seen in Table 3.

IMPLICATIONS

Implications for user

Often, a single app does not meet the needs/motivations of the user, and it is common for many participants to use multiple tracking apps. In a former research study, participants report that ‘some [apps] have features for health and some have features for fertility planning. . . to make the most, I have used various apps at the same time and entered data into them twice’ [2^o]. In other cases, women are interested in aligning their cycle data to health and activity data from wearables and other health tracking apps they may be utilizing [2^o]. Differences in terms of which symptoms are collected, and how,

Table 3. Varying granularity with which symptoms are recorded by different apps

Weight	Indicate weight Gain/loss	My Calendar, Period Tracker, Period Calendar, Maya
	Input numerical value	Clue, Glow, My Calendar, Period Tracker, Period Diary, Ovia, Cycles, Maya, Ladytimer, MyDays X, Fertility Friend, Woman Log
	Input numerical value and time of measurement	Life
Bleeding	Presence/absence of symptom	My Calendar, Period Tracker, Period Calendar, Fertility Friend
	Light, medium, heavy, spotting	Clue, Glow, Eve, Life, Ovia, Spot On, Ladytimer, Woman Log
	No flow, had flow, light, medium, heavy	Cycles, Apple Health, Pepapp, Period Diary
Spotting	Presence/absence of symptom	My Calendar, Period Tracker, Period Calendar, Cycles, Apple Health, Maya
	Indicate range (0–100%)	My Days X
	Indicate range (light, medium, heavy)	Glow, Period Tracker, Menstrual Period Tracker, Ladytimer, Woman Log
	Qualify color	Ovia
Cervical firmness	Presence/absence of symptom	My Calendar, Period Tracker Period Calendar
	Firm, medium, soft	Glow, Period Tracker, Life, Ovia, My Days X, Fertility Friend
Cervical openness	Presence/absence of symptom	My Calendar, Period Tracker Period Calendar
	Closed, medium, open, none	Glow, Period Tracker, Life, Ovia, My Days C, Fertility Friend
Cervical position	Presence/absence of symptom	My Calendar
	Low, medium, heavy	Glow, Period Tracker, Life, Ovia, My Days, Fertility Friend
Mucus quality	Presence/absence of symptom	My Calendar, Period Tracker Period Calendar, Maya
	Dry, sticky, watery, raw egg white, creamy	Clue, Glow, Period Tracker, Life, Ovia, Menstrual Period Tracker, Cycles, Apple Health, Maya, Ladytimer, My Days X, Fertility Friend
Cramps	Presence/absence of symptom	Clue, My Calendar, Period Tracker, Period Calendar, Period Diary, Ovia, Eve, Life, Cycles, Spot On, Flo, Apple Health, Pepapp, Fertility Friend
	Indicate severity	Glow, Period Tracker, Ladytimer, Woman Log
Headache	Presence/absence of symptom	Clue, My Calendar, Period Tracker Period Calendar, Eve, Life, Ovia, Cycles, Spot On, Apple Health, Maya, Pepapp, Fertility Friend, Flo
	Indicate severity	Glow, Period Tracker, Menstrual Period Tracker, Ladytimer, My Days X, Woman Log
Migraine	Presence/absence of symptom	Ovia, My Calendar, Period Tracker Period Calendar, Life
	Indicate severity	Glow, Ladytimer, My Days, Woman Log
Backache	Presence/absence of symptom	My Calendar, Period Tracker, Period Calendar, Period Diary, Eve, Life, Ovia, Cycles, Spot On, Cycle Tracker, Apple Health, Fertility Friend, Flo
	Indicate severity	Glow, Period Tracker, Ladytimer, My Days, Woman Log
Muscle/joint pain	Presence/absence of symptom	My Calendar, Period Tracker Period Calendar, Ovia, Pepapp
	indicate severity	Period Tracker, Ladytimer, My Days, Woman Log
Hot flashes	Presence/absence of symptom	Ovia, My Calendar, Period Tracker Period Calendar, Cycles, Apple Health
	Indicate severity	Glow, Ladytimer, Woman Log
Chills	Presence/absence of symptom	Ovia, My Calendar, Apple Health
	Indicate severity	Ladytimer, Woman Log
Pelvic pain	Presence/absence of symptom	Period Tracker Period Calendar, Apple Health
	Indicate severity	Glow
	left, right pelvic area	Ovia
Constipation	Presence/absence of symptom	Clue, My Calendar, Period Tracker, Period Calendar, Period Diary, Eve, Ovia, Spot On, Cycle Tracker, Apple Health, Maya, Pepapp, Fertility Friend, Flo
	Indicate severity	Glow, Period Tracker, Ladytimer, My Days, Woman Log

Table 3 (Continued)

Diarrhea	Presence/absence of symptom	Clue, My Calendar, Period Tracker Period Calendar, Eve, Ovia, Spot On, Apple Health, Maya, Pepapp, Fertility Friend, Flo
	Indicate severity	Glow, Ladytimer, Woman Log
Bloating	Presence/absence of symptom	Clue, My Calendar, Period Tracker Period Calendar, Eve, Life, Cycles, Spot On, Apple Health, Maya, Pepapp, Fertility Friend, Flo
	Indicate severity	Glow, Ladytimer, Woman Log
Skin/acne	Presence/absence of acne	My Calendar, Period Tracker, Period Calendar, Period Diary, Eve, Life, Ovia, Cycles, Spot On, Cycle Tracker, Apple Health, Pepapp, My Days, Fertility Friend, Woman Log, Flo
	Severity of acne	Glow, Period Tracker, Ladytimer
	Skin condition –oily, dry, acne	Clue, Woman Log
Mood	Presence/absence of moodiness	My Calendar, Eve, Apple Health
	0–5 states	My Calendar, Period Tracker Period Calendar, Cycle Tracker, Pepapp, Fertility Friend
	5+ states	Glow, Period Tracker, Life, Ovia, Cycles, Maya, Ladytimer
Sleep	Presence/absence of insomnia	My Calendar, Period Tracker, Period Tracker Period Calendar, Ovia, Cycles, Maya, Fertility Friend, Flo
	Severity of disturbed sleep	Glow, Ladytimer, My Days, Woman Log
	Time slept (quantified numerically)	Clue, Glow, Period Tracker, Ovia
	Descriptions of quality of sleep	Life, Spot On
Sex drive	Presence/absence of increased/decreased sex drive	Life, Ovia, My Days, Fertility Friend, Flo
	Qualify sex drive on a spectrum	Glow, Eve, Ladytimer
Cravings	Presence/absence of cravings	My Calendar, Period Tracker Period Calendar, Eve, Cycles, Apple Health, Maya, Flo
	Increase/decrease in appetite	Ovia, My Days, Fertility Friend
	Qualify appetite along a scale	Glow
	Severity of cravings	Period Tracker, Ladytimer, Woman Log
	Types of cravings (sweet, salty, carbs, chocolate)	Clue, Eve, Spot On, My Days, Woman Log
Exercise	Engaged in exercise	Spot On, Fertility Friend
	Choose from activities (running, yoga, biking, swimming)	Ovia, Clue, Eve
	Minutes of exercise (quantified numerically)	Ovia, Glow, Period Tracker, Ladytimer

The manner in which different symptoms are recorded by different apps can be seen in Table 3. It is important to note that this is representative of manually inputted data only, and not data that is ingested from the Apple Health Cycle Tracker or external wearables.

across different apps, however, makes aggregating data collected from different apps a nearly impossible task.

No data can be shared between two different menstrual health tracking apps. Although the addition of Apple's Cycle Tracker presents the opportunity for all relevant health data to be stored in one central location of the apps reviewed only seven (Clue, Glow, Period Tracker, Eve, Life, MPT, Flo, Woman Log, Cycles) provide the user the ability to sync data with the native Apple Health app. Two (Period Tracker and Ovia) can import data collected from Apple Health or a Fitbit. In addition, the Ovia app can share data with other Ovia-owned apps specific to pregnancy planning.

The data that can be read and written between the Apple Health and most menstrual health tracking apps reviewed in the article is basal body temperature, cervical mucus quality, menstruation, ovulation test result, sexual activity, and spotting. Symptom-specific information, multiple metrics of which overlap between tracking apps and Apple Health, however, cannot be shared. A complete description of data that can be shared among the different apps is represented in Fig. 5.

Eight of the apps reviewed (My Calendar, Spot On, Period Diary, Maya, Ladytimer, Pepapp, MyDay, Fertility Friend) do not currently provide data integration capabilities, the lack of which can mean that users need to manually enter data into multiple apps [4].

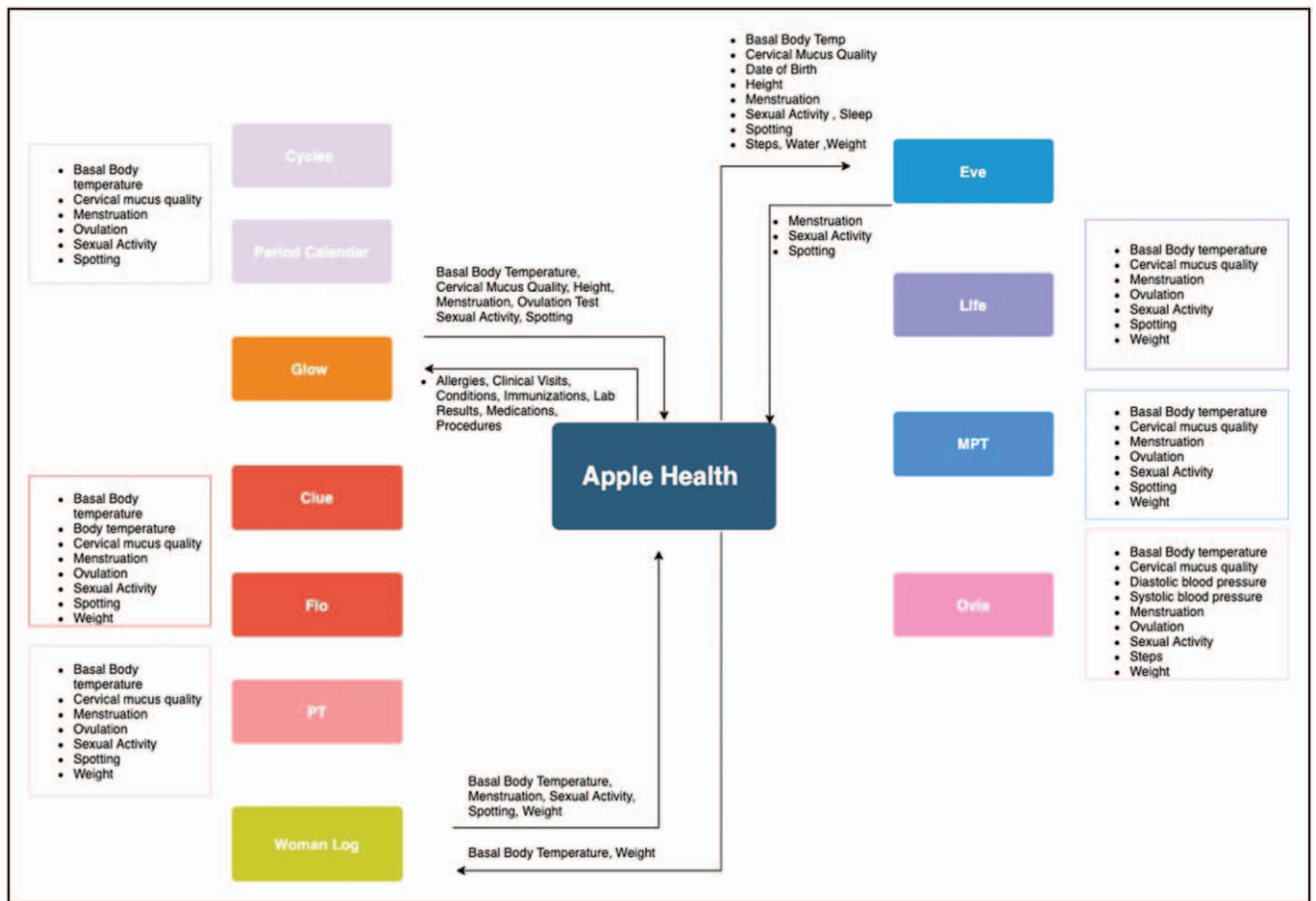


FIGURE 5. Data integration with Apple Health-User Perspective. Most of the apps reviewed can integrate data with Apple Health. For apps where the data that can be read and written coincides, the metrics have been listed in color-coded text boxes. Apps where the data that can be read from Apple Health differs from the app that can be written to Apple Health, arrows indicate the directionality of the information shared. The app Eve, for instance, can ingest basal body temperature, cervical mucus quality, date of birth, height, menstruation, sexual activity, sleep, spotting, step, water, and weight from Apple Health. The only data that can be written to Apple Health, however, is menstruation, sexual activity, and sleep.

Implications for researchers

The variation in infrastructure that is used to track symptoms means that the data housed in numerous health tracking apps must be processed, and may introduce bias, prior to data analysis. The variance in key terminology and lack of definitions means that even within a specific app, researchers cannot ensure that each category holds the same meaning to each user, and app-specific conclusions can be questioned [9]. Further complicating the issue, users are not consistent and thorough in their tracking practices. Most users report changes in the number and categories of symptoms recorded over the course of their interaction with an app [10]. Numerous studies, including some specific to premenstrual syndrome (PMS) indicate that the presentation of symptoms and having knowledge of symptoms can lead to initial exacerbated symptom recording. Once the novelty of the app has decreased, and with

increased self-awareness, however, users no longer feel the need to record as frequently [2,10]. Researcher must, therefore, evaluate the quality of the data collected and strategize to eliminate self-tracking artifacts [11].

There is, however, value in prospective research studies that investigate associations of specific phenotypes to menstrual cycle characteristics. The prospective design of such studies allows for greater control of user demographics, to ensure representation from all populations of interest and helps monitor user engagement. Significant knowledge gaps that remain can be addressed with such an approach. Studies of the relationship between menstrual patterns and symptomatic variables are limited –recent work has explored these associations using self-tracked data but over a limited set of symptoms and without accounting for age or birth control usage [9]. There are even fewer studies

examining the relationship between emotional symptoms and the experience of physical pain specifically during a menstrual cycle, despite there being significant evidence that this link exists for other health conditions [12,13]. Also lacking are studies that consider individuals who experience irregular periods, are struggling to conceive, and do not conform to a gender binary [14].

Implications for clinicians

Menstrual health apps allow users to share accurate details with their medical providers [2^o]. Digital self-tracking compared with paper-based tracking or memory-reliant surveys supplies more cohesive and accurate records [14]. Four of the apps reviewed (Glow, Period Calendar, Life, $n = 4$) even offer means to summarize and export data that can be shared with a user's healthcare provider.

Although clinicians express interest in promoting the use of health apps, they are presented with the challenge of recommending an app to their patients. Given that apps differ greatly in the number of symptoms they record and the specificity to which the symptoms are tracked, there is a need to achieve a tradeoff between the two metrics. Presenting a user with apps that record numerous symptoms with great granularity may lead to user fatigue and diminished data completeness, whereas using an app that records only a few symptoms with great granularity can lead to misinformed conclusions [15]. This is further complicated by the rates at which current apps are updated and new apps are introduced. The pace of such changes exceeds the rate at which research around the apps can occur [16].

Understanding where there is a need to record symptoms with greater specificity such that the use of an app can provide interesting insights into menstrual health and rhythmicity is imperative. Nonetheless, the variable of interest is specific to patient needs. With steadily increasing numbers of apps specific to menstrual health conditions, such as Phendo, and those targeted at various reproductive stages, menarche onset and menopause, the task of choosing an app may become more streamlined. App integration into clinical routines, meanwhile, must use an intentional approach. Healthcare providers can develop app selection tools or adopt existing ones and cultivate a knowledge base by dedicating medical team members to the task [17].

Clinical relevance of symptoms

Establishing clinical associations and relevance of the symptoms presented by the apps; however,

remains an issue, which can only be addressed through collaborations between clinical and epidemiological researchers. Establishing the relevance of symptoms is complicated by the limited number of clinical guidelines, and lack of adequate knowledge governing the management of menstrual health disorders, such as endometriosis and polycystic ovarian syndrome (PCOS) [18,19]. The physical symptoms collected by apps, particularly gastrointestinal symptoms including abdominal pain, diarrhea, constipation, nausea, and vomiting, are consistent with symptoms that have been used to discriminate PMS and menstrual experiences [19]. Increasingly, however, there are attempts to characterize phenotypic indications of menstrual health over complete lifecycles and distinguish among menstrual health disorders. Phendo, an app designed to phenotype endometriosis, found that participants with self-diagnosed endometriosis report pelvic and lower back pain, cramping, gastrointestinal symptoms, menorrhagia, spotting outside of the period, painful sex, avoidance of sex, clotting, fatigue, mental foggy, and headache [20,21]. A similar research study conducted to diagnose PCOS assigned significance to variables, such as irregular menstrual cycles, hirsutism, alopecia, and acne in addition to the Rotterdam criteria – androgen excess, menstrual irregularity, polycystic ovary morphology on ultrasound [22,23]. Apps built primarily for menopause symptom tracking, mySysters and Menopause View, often allowing tracking bladder pain, body odor, brittle nails, burning vagina, dry eyes, fecal incontinence, heart palpitations, metallic taste, tingling in extremities, tinnitus, and a deepening voice.

Condition-specific apps are not advertised as diagnostic tools, instead claiming to 'track, manage, and understand' the conditions. The low granularity with which apps are currently collecting symptoms does not allow apps to realize their diagnostic potential. Pain, for instance, is commonly captured as merely present or absent in menstrual health tracking apps or is ranked along a vague mild–severe scale. Medical-grade assessments of pain, on the other hand, involve documenting onset, location, intensity, duration, exacerbating/relieving factors, and whether the pain radiates.

To convert data collected from apps to digital diagnostic tools, clinicians must work towards establishing diagnostic associations between symptoms and menstrual health conditions. This process could leverage data already available through health-tracking apps and further controlled prospective studies. Patient experiences and interest can then serve to reinforce the associations identified [24]. Data scientists and researchers will be

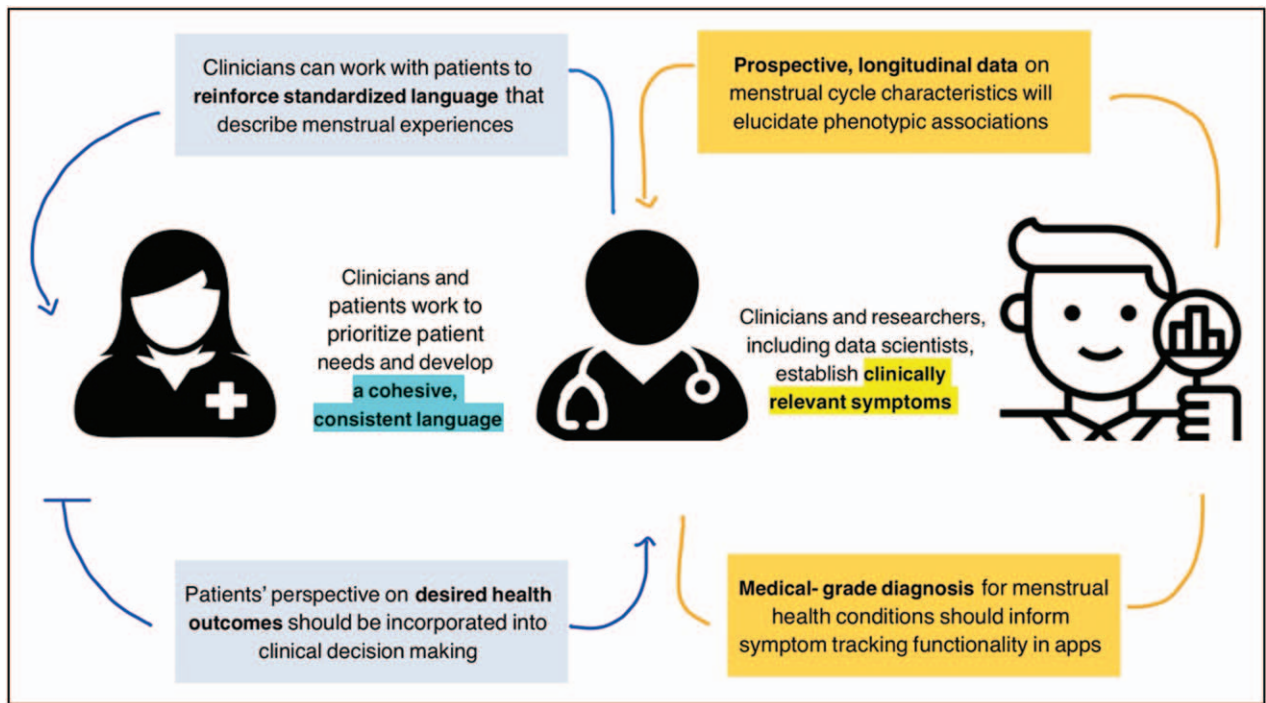


FIGURE 6. Clinicians lie at the heart of collaborative efforts attempting to transform menstrual health tracking apps into diagnostic tools. Clinicians must first work with patients to ensure that their needs are being represented. Patient experiences and needs, alongside a clinician’s expertise should inform the symptom tracking functionality in apps. Data scientists can then ensure that the data is collected in a manner that is harmonizable. The process is self-reinforcing such that phenotypic associations that emerge from data analysis will facilitate establishing the clinical relevance of symptoms, which can then be prioritized in health app design.

involved in both supporting clinicians as data is leveraged to create clinical guidelines, and in redesigning apps such that the data is collected to realize its diagnostic value (Fig. 6).

CONCLUSION

Currently, menstrual health tracking apps do not track symptoms in a manner conducive to research efforts. There is a lack of standardized language in describing and characterizing menstrual experiences. Moreover, different symptoms are collected with varying degrees of specificity. Nonetheless, none of the general or condition-specific menstrual tracking apps collect symptoms in a manner that may classify as clinical grade. Collecting data in a more effective, diagnosis-friendly manner requires establishing the clinical relevance of symptoms. In the process, clinicians can leverage patient experiences and data from existing and upcoming digital epidemiological studies. The guidelines established will then allow data scientists and developers to dictate data flow through menstrual health tracking apps such that they can be used as diagnostic tools.

Recommendations

Adopting a consistent vocabulary to describe relevant symptoms is a crucial step towards realizing the potential of data from menstrual health tracking apps. Given the importance of tracking bleeding days to most users, limiting the use of bleeding, spotting, flow, and discharge to specific cases would increase data usability. For instance, spotting can be used to describe bleeding between periods, whereas flow can be reserved for bleeding days and collected along a scale. Although qualifiers of cervical mucus quality should remain consistent, the terms used to qualify cervical position, openness, and firmness could be standardized. By further providing clear definitions of symptoms, and explanations of how a particular metric should be recorded, developers can mitigate some concerns related to variability in user interpretation and understanding.

Assigning particular metrics to specific type, or granularity, of recording would greatly simplify and expand the amount of data that could be synced across health tracking apps and wearables. For instance, amount of time spent exercising and sleeping and metrics, such as body weight and temperature could be collected, universally, as numerical

values. To quantify changes in behavioral variables, including appetite and sex drive, scales ranging from 'lower than usual, low, medium, high, higher than usual' could present a compromise between current practices that allow a user to either indicate +/- in behavior or rank the behavior along a scale that roughly corresponds to low, medium, and high. The standardized qualifiers of cervical state should be adopted by apps that currently ask users to qualify the metrics along a scale.

Although most physical symptoms can initially be collected at the present/absent level, apps could present more detailed recording of symptoms that re-occur for particular users. Period-related pain, specifically, could be collected in manner, which resembles pain assessment in clinical environments.

Although challenging, catalyzing the adoption of such recommendations could model past harmonization efforts by the NIH, such as the guidance developed for Environmental Influences on Child Health Outcomes data harmonization efforts. A symposium with representation from clinicians, epidemiological researchers, and app developers will help create a distributable collection of best practices. Although clinicians could provide insight into what data might be considered priority for harmonization, epidemiological researchers can suggest the most meaningful data collection format, and inviting comments and suggestions from app developers would ensure crucial stakeholder buy-in.

Limitations

One limitation of the review presented is that the apps analyzed were from the iTunes AppStore. Apps that are only available through Google Play were not included. Previous research indicates that users of Androids usually fall within a lower socioeconomic barrier than iPhone users, and it may be of interest to characterize any difference in symptom tracking functionality and behavior among the two user groups [25]. Furthermore, the apps were piloted through a single simulated cycle, which limits our ability to analyze the symptom trend reporting feature some apps presented. Lastly, we acknowledge that apps are frequently updated, and it cannot be guaranteed that the information presented is the most recent, accurate representation of what is being offered to users.

Acknowledgements

None.

Financial support and sponsorship

This research was supported in part by the intramural research program of the National Institutes of Health,

National Institute of Environmental Health Sciences under award number Z01ES103333.

Conflicts of interest

There are no conflicts of interest.

REFERENCES AND RECOMMENDED READING

Papers of particular interest, published within the annual period of review, have been highlighted as:

- of special interest
- of outstanding interest

1. Women's Health App Market Size Analysis Report 2021-2028 Women's Health App Market Size, Share & Trends Report Women's Health App Market Size, Share & Trends Analysis Report By Type (Fitness & Nutrition, Pregnancy Tracking & Postpartum Care, Menopause), B. Women's Health App Market Size Analysis Report 2021-2028. 2021.

2. Levy J, Romo-Avilés N. A good little tool to get to know yourself a bit better": a qualitative study on users' experiences of app-supported menstrual tracking in Europe. *BMC Public Health* 2019; 19:1213.

The article analyzes, in-depth, eight characteristics of app-based menstrual tracking, presenting identifying irregularities as a primary motivator amongst users. It further suggests that the most prominent tracked metrics are emotion, ovulation, pain, sexual activity, and vaginal discharge. The varied motivators and tracking behaviors led to both negative and positive patient experiences. The study shows that while certain patients were distressed over irregularities and wrongly associated symptoms with menstruation, others were able to use the data to educate themselves and productively share information with healthcare providers.

3. Lee J, Kim J. Can menstrual health apps selected based on users' needs change health-related factors? A double-blind randomized controlled trial. *J Am Med Inform Assoc* 2019; 26:655-666.

The study verifies that when using apps that best reflect their needs, amongst a population of menstruation experiencing dysmenorrhea and PMS-related characteristics, users record more often, report higher app satisfaction, and intend to recommend the app to others.

4. Epstein DA, Lee NB S Kang JH, *et al.* Examining menstrual tracking to inform the design of personal informatics tools. *Proc SIGCHI Conf Hum Factor Comput Syst* 2017; 2017:6876-6888.

5. Bull JR, Rowland SP, Schervitz EB, *et al.* Real-world menstrual cycle characteristics of more than 600,000 menstrual cycles. *Npj Digital Med* 2019; 2. <https://doi.org/10.1038/s41746-019-0152-7>.

Using a digital epidemiology approach, the article analyzes real-world menstrual cycle data from a health tracking app, to improve our understanding of cycle physiology and characteristics. The analysis reveals great variability in cycle length, caused by differences in follicular phase length, across women of different ages and BMIs ultimately challenging assumptions surrounding a 'normal' cycle.

6. Li KU. Characterizing physiological and symptomatic variation in menstrual cycles using self-tracked mobile-health data. *NPJ Digit Med* 2020; 3:79.

The article aimed to first determine variability in cycle length characteristics, and then investigate differences in symptomatic behavior between high-variability and low-variability menstruators using data from a menstrual health tracking app, Clue. The study found that period flow, emotional state, and experienced pain were amongst the most frequently recorded categories across both groups. However, women located at different ends of the variability spectrum did indeed exhibit different symptom patterns. Women with high-variable cycle length characteristics exhibited more heterogeneous period tracking and were unlikely to report heavy periods.

7. Soumpasis I, Grace B, Johnson S. Real-life insights on menstrual cycles and ovulation using big data. *Human Reproduct Open* 2020; 2020:hoaa011.

The study leverages data from a connected home ovulation test system, which combines the use of urinary ovulation tests and an accessory app, to reveal differences between a women's perceived cycle length and actual cycle length and suggest that miscalculations of one's fertile window contribute significantly to problems with conception.

8. Critchley HOD, Babayev E, Bulun SE, *et al.* Menstruation: science and society. *Am J Obstet Gynecol* 2020; 223:624-664.

9. Schantz JS, Fernandez CS, Jukic AM. Menstrual cycle tracking applications and the potential for Epidemiological research: A comprehensive review of the literature. *Current Epidemiology Reports* 2021; 8:9-19.

The article evaluates the potential of menstrual cycle tracking apps for epidemiological studies. Identified advantages include access to diverse population level data and circumventing traditional 'recruitment' efforts. Missing data and its implications on interpretation, however, is cited as a major concern. Moreover, there is a potential for selection bias. For instance, the study found a tendency for app users included in primary research studies to be of the white race or European residence. Similarly, studies evaluating the accuracy of fertile window reporting by apps may be unintentionally overrepresenting views of users struggling to conceive.

10. MacKrell K, Groom KM, Petrie KJ. The effect of symptom-tracking apps on symptom reporting. *Br J Health Psychol* 2020; 25:1074–1085.
11. Li K, Urteaga I, Shea A, *et al.* A predictive model for next cycle start date that accounts for adherence in menstrual self-tracking. *J Am Med Inform Assoc* 2021. <https://doi.org/10.1093/jamia/ocab182>.
12. Pierson EA, Leskovec J. Modeling individual cyclic variation in human behavior. *Proc Int World Wide Web Conf* 2018; 2018:107–116.
13. Pierson E, Althoff T, Thomas D, *et al.* Daily, weekly, seasonal and menstrual cycles IN women's Mood, behaviour and vital signs. *Nat Hum Behav* 2021; 5:716–725.
14. Laganà AS, La Rosa VL, Rapisarda AMC, *et al.* Anxiety and depression in patients with endometriosis: impact and management challenges. *Int J Womens Health* 2017; 9:323–330.
15. Moglia ML, Nguyen HV, Chyjek K, *et al.* Evaluation of smartphone menstrual cycle tracking applications using an adapted APPLICATIONS scoring system. *Obstet Gynecol* 2016; 127:1153–1160.
16. Choe EK, Lee NB, Lee B, *et al.* Understanding quantified-selfers' practices in collecting and exploring personal data. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. 2014. pp. 1143–1152.
17. As-Sanie S, Black R, Giudice LC, *et al.* Assessing research gaps and unmet needs in endometriosis. *American Journal of Obstetrics and Gynecology* 2019; 221:86–94.
18. Carmina E, Lobo RA. Polycystic ovary syndrome (PCOS): arguably the most common endocrinopathy is associated with significant morbidity in women. *J Clin Endocrinol Metab* 1999; 84:1897–1899.
19. Fiordelli M, Diviani N, Schulz PJ. Mapping mHealth research: a decade of evolution. *J Med Internet Res* 2013; 15:e95.
20. McCartney PR. Nursing practice with menstrual and fertility mobile apps. *MCN Am J Matern Child Nurs* 2016; 41:61.
21. Bernstein MT, Graff LA, Avery L, *et al.* Gastrointestinal symptoms before and during menses in healthy women. *BMC Women's Health* 2014; 14:14.
22. Urteaga I, McKillop M, Lipsky-Gorman S, Elhadad N. Phenotyping endometriosis through mixed membership models of self-tracking data. *Proc Machine Learning Healthc* 2018.
23. Urteaga I, McKillop M, Elhadad N. Learning endometriosis phenotypes from patient-generated data. *NPJ Digit Med* 2020; 3:88.
24. Bretschneider RA. A goal and context-driven approach in mobile PERIOD tracking applications. *Universal Access in Human-Computer Interaction. Access to Learning, Health and Well-Being*. 2015: 279–287. https://doi.org/10.1007/978-3-319-20684-4_27.
25. Jamalova M, Milán C. The comparative study of the relationship between smartphone choice and socio-economic indicators. *Int J Market Studies* 2019; 11:11.