International Journal of Advance and Innovative Research
Volume 13, Issue 1 (XIII): January – March 2026
ISSN 2394 - 7780

A DATA-DRIVEN APPROACH FOR IDENTIFYING GAPS IN CYCLE SYNCING AWARENESS AMONG WOMEN

Bhumika Dube[1]*, Sumitkumar Tripathi[2] and Kanojia Mahendra[3]
[1]Department of Information Technology, Sheth L.U. Jhaveri and Sir M.V. College, India, dubebhumika.it@gmail.com
[2]Department of Information Technology, Sheth L.U. Jhaveri and Sir M.V. College, India, sumit11.tripathi@gmail.com
[3]Department of Computer Science, Sheth. L.U.J. and Sir M.V. College, India, kgkmahendra@gmail.com
Corresponding author: Bhumika Dube, dubebhumika.it@gmail.com

ABSTRACT
Menstrual health literacy is a crucial determinant of women's physical health and overall well-being and daily functioning. In recent years, cycle syncing has emerged as a structured lifestyle approach that encourages aligning lifestyle practices such as diet, work intensity, exercise, and self-care with the distinct phases of the menstrual cycle and follow them regularly in daily life. Despite its potential benefits, primary empirical findings indicate that women remain insufficiently aware of cycle-syncing practices, and their consistent application in daily life is still underexplored. The present study addresses this gap by examining women's literacy regarding the cycle-syncing concept and assessing whether awareness translates into regular practice or not. The primary objective is to evaluate awareness levels, behavioral adoption, and menstrual health practices related to cycle tracking and irregularities. A descriptive, survey-based methodology was applied using a structured Google Form distributed online to female participants. Quantitative data collected from female respondents from which accurate female responses were analyzed to identify patterns in awareness, lifestyle alignment, cycle tracking behavior, and perceptions of menstrual irregularities such as delayed or missed periods. The findings indicate that while a majority of respondents demonstrate awareness of menstrual cycle phases and the concept of cycle syncing, but still nearly half of them do not actively follow cycle-synced practices in their routine lives. Additionally, a significant proportion reported experiencing irregular or delayed cycles, often without systematic tracking or timely attention. The study highlights a clear gap between conceptual awareness and practical implementation, underscoring the need for targeted educational initiatives and digital health interventions to enhance menstrual health literacy and promote sustained, cycle-aware lifestyle practices among women.

Keywords: *Menstrual health literacy, cycle syncing, reproductive health behavior, lifestyle alignment, cycle tracking practices, menstrual irregularities, empirical survey study*

1. INTRODUCTION
Every woman goes through menstruation once a month, a natural biological process that often involves bleeding, abdominal cramps, fatigue, mood changes, and other physical or emotional discomforts. Beyond these commonly experienced symptoms, menstruation reflects an underlying cycle of hormonal and reproductive changes that play a crucial role in women's overall health and well-being. The menstrual cycle typically spans about 28 days, although durations ranging from 21 to 35 days are considered normal, and is divided into four distinct phases: menstrual, follicular, ovulatory, and luteal. The menstrual phase marks the onset of the cycle and generally lasts 3-7 days, during which uterine shedding occurs. This is followed by the follicular phase, approximately between days 8 and 13, when rising estrogen levels stimulate egg development and prepare the body for ovulation. Ovulation usually occurs around day 14 with the release of an egg from the ovary. The cycle concludes with the luteal phase, typically spanning days 15 to 28 (or up to 35 in longer cycles), during which progesterone supports potential implantation or leads to the initiation of the next cycle in the absence of pregnancy (Maharashtra State Bureau of Textbook Production and Curriculum Research, 2022). Research shows that many menstruating individuals lack comprehensive knowledge of their cycles, which can impact symptom recognition, management of disorders such as Premenstrual Syndrome (PMS), and informed health choices across the reproductive lifespan (Cunningham et al., 2024; Goddard et al., 2025). While mobile health technologies and targeted interventions have shown promise in improving menstrual and reproductive health awareness, knowledge levels remain low in diverse settings without structured education (Baird, 2022; Goddard et al., 2025).

Poor awareness regarding such topics may lead to the neglect of early warning signs associated with disorders such as polycystic ovary syndrome, endometriosis, anemia, and hormonal imbalances, which can adversely affect long-term physical and mental health if left unaddressed (Baird, 2022; Gopalan et al., 2024; Hennegan et al., 2019; Itriyeva, 2022). Menstrual cycle tracking and self-monitoring are increasingly popular behaviors that may support greater awareness of cycle phases, variability, and irregularities, but the motivations and consistency of such practices vary widely and underlying behavioral adoption patterns are not yet fully characterized in recent empirical work (Stujenske, 2023). Despite the growth of digital tools and cycle tracking apps, there is limited evidence on whether awareness of menstrual patterns translates into sustained behavior change, especially in lifestyle alignment strategies like cycle syncing. Moreover, reviews of menstrual symptoms and activity suggest that individual variability in cycle experiences can affect daily functioning and health behaviors, highlighting the importance of personalized literacy and management skills (Gopalan et al., 2024). The present study aims to address this gap by examining women's level of literacy related to the cycle syncing concept and evaluating whether awareness influences regular implementation in daily life.

The objective of this study was to assess the level of awareness and real-life adoption of the cycle-syncing concept among women using primary data collected through a structured Google Form questionnaire. The study aimed to establish baseline menstrual cycle literacy, evaluate participants understanding of menstrual cycle phases, and examine their familiarity with cycle syncing and exposure to menstrual health information through educational resources, digital platforms, and social media, which play a critical role in shaping menstrual health awareness in the digital era (Schantz et al., 2021). In addition, the study investigated the extent to which participants implement cycle-synced lifestyle practices such as aligning physical activity, dietary intake, rest patterns, and self-care routines with different phases of the menstrual cycle and explored the perceived impact of these practices on menstrual symptom management and overall well-being (Chen et al., 2023). Furthermore, the study sought to compare conceptual awareness with actual behavioral implementation in order to identify gaps between knowledge and sustained adoption of cycle-syncing practices, as well as to examine key barriers influencing inconsistent or non-adoption, including informational limitations, irregular menstrual cycles, cultural stigma, and lack of professional guidance (Schantz et al., 2021; Chen et al., 2023).

In this study, awareness of cycle syncing is treated as the independent variable (IV), while real-life adoption of cycle-syncing practices is considered the dependent variable (DV). Let, X_A denote awareness of the cycle-syncing concept & X_P denote real-life adoption of cycle-syncing practices. Both variables are operationalized as binary outcomes, where a value of 1 indicates presence (awareness or adoption) and 0 indicates absence. Accordingly, the variables are modeled as Bernoulli random variables, a standard approach for analyzing dichotomous behavioral and awareness-based outcomes in health and social science research.

$$X_A \sim Bernoulli(p_A), X_P \sim Bernoulli(p_P) \quad (1)$$

In equation 1, p_A represents the probability of awareness of cycle syncing, and p_P represents the probability of adopting cycle-syncing practices in real life. The use of Bernoulli modeling and conditional probability provides a statistically robust framework for examining behavioral adoption influenced by awareness, as supported in foundational probability and applied biostatistics literature (Ross, 2019). Bernoulli modeling forms the probabilistic basis, while conditional probability is later used to test the effect of awareness on adoption. The null hypothesis posits that there is no significant relationship between awareness of the cycle-syncing concept and real-life adoption of cycle-syncing practices among women. In other words, the probability of adopting cycle-syncing behaviors remains unchanged regardless of whether an individual is aware of the concept.

$$H_{0_1}: P(X_P = 1 \mid X_A = 1) = P(X_P = 1) \quad (2)$$

The alternative hypothesis proposes that awareness of the cycle-syncing concept significantly increases the likelihood of its adoption in real-life practices. This hypothesis assumes that informed individuals are more likely to translate menstrual health knowledge into actionable lifestyle behaviors.

$$H_{1_1}: P(X_P = 1 \mid X_A = 1) > P(X_P = 1) \quad (3)$$

The equation 2 defines the Null Hypothesis, it states that the probability of adopting cycle-syncing practices (X_p = 1) remains the same regardless of whether an individual is aware of the cycle-syncing concept (X_a = 1). Mathematically, this formulation implies no statistically significant association between awareness and adoption. Awareness does not alter behavior any observed adoption occurs independently of knowledge about cycle syncing. The equation 3 defines the Alternative Hypothesis, it proposes that awareness of the cycle-syncing concept increases the likelihood of its adoption in real-life practices. Formally, this inequality captures the assumption that informed individuals are more likely to translate menstrual health knowledge into actionable lifestyle behaviors, indicating a positive and meaningful effect of awareness on adoption. Together, equations 2 and 3 precisely operationalize the research hypothesis by converting a conceptual relationship (awareness vs. adoption) into testable probabilistic statements, enabling rigorous statistical validation.

2. LITERATURE REVIEW

The study by Francois (2024) addresses the emerging concept of cycle syncing by examining how aligning nutrition, exercise, and structured lifestyle practices with menstrual cycle phases may optimize women's quality

 ISSN 2394 - 7780

of life and their well-being. Through a comprehensive literature review and observational analysis of credentialed social media sources, the study highlights that hormonal fluctuations across the menstrual cycle significantly influence mood, energy, cognition, and physical performance. While the findings suggest potential benefits of phase-specific lifestyle adjustments, the study remains largely conceptual and observational in nature, lacking primary survey or longitudinal data on women's real-life adherence to cycle-syncing practices. This limitation underscores the need for empirical research examining women's awareness, consistency of adoption, and practical challenges associated with implementing cycle-based lifestyle strategies in everyday contexts. The study by Saini and Dutta (2025) addresses the underexplored role of menstrual cycle rhythms within psychological and therapeutic settings by introducing hormonal awareness into psychotherapy practice. Using a biopsychosocial perspective, the authors highlight how hormonal fluctuations across menstrual phases influence mood, emotional regulation, motivation, and behavior. While the paper emphasizes the clinical relevance of cycle awareness for tailoring therapeutic interventions, it remains conceptual and viewpoint-based, without primary behavioral or survey data. The study does not examine whether menstrual or cycle-syncing awareness translates into structured lifestyle or health practices beyond therapy rooms. This limitation highlights a gap in empirical research linking menstrual awareness to real-world behavioral adoption and daily health management.

The study by Pfender *et al.* (2025) investigates the proliferation of cycle-syncing content on TikTok and critically examines the evidence behind messaging related to aligning diet, exercise, and lifestyle with the menstrual cycle. By analyzing dozens of top-ranked videos, the researchers found that most content creators promote cycle syncing with general recommendations for nutrition and physical activity mapped to cycle phases, yet very few cite scientific evidence to support these claims. The study highlights how social media oversimplifies complex biological literature, amplifying anecdotal benefits without robust empirical backing and scientific knowledge. This research underscores the gap between popular cycle-syncing trends and validated scientific knowledge, pointing to the need for expert-driven, evidence-based health communication strategies around menstrual health. The article published on Signos (2024) discusses the concept of cycle syncing and its proposed benefits for metabolic and hormonal health by aligning lifestyle behaviors with menstrual cycle phases. It outlines how nutritional intake, physical activity, and daily habits may be tailored to specific phases to support energy levels, appetite regulation, and metabolic efficiency. While the piece provides a useful overview of theoretical cycle-based lifestyle adjustments, it is primarily informational and does not present original empirical data or experimental findings. Consequently, this source highlights the conceptual appeal of cycle syncing but underscores the absence of rigorous, data-driven evidence confirming its effectiveness in real-world health outcomes, reinforcing the need for primary research that examines actual awareness, adoption, and measurable impacts of cycle-based practices.

The study by Itriyeva (2022) provides a comprehensive clinical overview of the normal menstrual cycle, emphasizing the coordinated functioning of the hypothalamic pituitary ovarian axis in regulating follicular development, ovulation, and luteal activity. The paper explains how hormonal feedback mechanisms govern cycle regularity and highlights that deviations such as delayed, irregular, or absent menstruation may indicate underlying health concerns. While the study offers strong physiological and clinical insights into menstrual irregularities, it does not explore women's awareness of these variations or the role of lifestyle-based strategies in cycle regulation. This limitation underscores the need for research linking menstrual physiology with menstrual literacy, self-monitoring practices, and behavioral interventions. The student research poster by Kapitula *et al.* (2023) investigates gaps in menstrual cycle knowledge and educational experiences among women. Through a survey examining what women know or believe they know about menstrual physiology and where they acquired that understanding, the study highlights that formal education often focuses only on basic reproductive concepts, leaving significant gaps in detailed menstrual cycle literacy. Participants reported limited understanding of cycle timing, phase variability, and biological mechanisms beyond fertility, suggesting that educational curricula may inadequately prepare individuals to interpret cycle patterns or health signals. While the poster provides important preliminary insights, it is exploratory in nature and does not extend to behavior or lifestyle applications of menstrual awareness, indicating a need for more robust empirical research in this area.

Based on the reviewed literature, menstrual cycle awareness among women is gradually increasing. However, this awareness often does not translate into consistent practical application in daily life with a scientific approach. While many women possess basic knowledge of menstrual physiology and hormonal changes, this understanding frequently remains theoretical and is not systematically integrated into lifestyle decisions such as diet, exercise, self-care, or health monitoring. Although concepts like cycle syncing have gained visibility through digital and social media platforms, existing research is largely descriptive or observational, with limited empirical evidence on real-world adoption and sustained outcomes. Persistent gaps in menstrual literacy, combined with cultural stigma, misinformation, and the lack of structured education, further hinder the transition from awareness to practice.

Collectively, these findings highlight a clear gap between knowledge and behavior, emphasizing the need for primary, data-driven research to examine actual implementation of cycle-based practices and their influence on menstrual health management. Nothing actually explains the concept of cycle-syncing and how following it regularly can improvise the women's entire life and well-being as well as their menstrual health.

3. PROPOSED WORK

The study engaged a quantitative, survey-based data collection method to examine awareness and adoption of cycle-syncing practices among women in their daily life. Primary data were collected using a structured online questionnaire designed and distributed through Google Forms, which facilitated efficient data gathering and management. The questionnaire consisted of 14 questions, of which 13th were close-ended Yes/No to capture measurable patterns of awareness, understanding, and behavioural adoption, while the 14th question was open-ended, allowing respondents to express their opinions, experiences, and personal thoughts regarding cycle-syncing and menstrual health. A total of 124 valid responses were received, providing a substantial sample for analysing audience engagement with the cycle-syncing concept. The combination of closed-ended and open-ended questions enabled both statistical assessment and qualitative insight into how women interpret menstrual health information and translate it into everyday practices. Google Forms was selected for its accessibility, ease of distribution, and ability to systematically record responses, thereby minimizing manual data handling errors. Participation in the study was entirely voluntary, and respondents were informed of the study's purpose prior to participation. To ensure ethical compliance and respondent comfort, no personally identifiable information was collected, and all responses were recorded. This approach encouraged honest engagement and authentic audience responses, aligning with the study's objective of understanding real perceptions and behavioural tendencies related to cycle-syncing.

A total of 121 valid responses out of 124 were selected for further data processing and analysis in the present study. All collected responses were systematically reviewed during the preprocessing stage to assess their completeness, consistency, and relevance to the research objectives. Responses that were incomplete, duplicated, or contained inconsistent or unreliable entries were excluded to ensure the accuracy and reliability of the dataset. The finalized dataset comprised 121 complete and valid responses from female participants who had voluntarily participated in the survey and provided full information across all questionnaire items. These responses offered sufficient and consistent data on key variables related to awareness of the cycle-syncing concept and its adoption in real-life practices, making them suitable for statistical analysis and interpretation.

The dataset was further refined through a structured data filtering process to enable focused subgroup-level analysis. Data filtering was conducted to systematically distinguish participants based on levels of awareness and real-life adoption of cycle-syncing practices while preserving the integrity and size of the original dataset. This step was essential for transforming categorical survey responses into analyzable numerical formats and ensuring compatibility with subsequent probabilistic and statistical analyses. Given that the questionnaire primarily consisted of categorical responses, an initial categorical-to-numerical conversion was performed prior to filtering. Such encoding is a standard preprocessing step in behavioral and health survey research, as it allows qualitative responses to be incorporated into quantitative analytical frameworks (Geron, 2019; Witten et al., 2016).

Binary and ordinal encoding schemes were applied depending on the measurement scale of each variable. For responses related to awareness of the cycle-syncing concept, the variable X was used to represent awareness status. These variables capture whether a participant is aware of cycle syncing. Responses were numerically encoded using a binary scheme, where a value of 1 indicates "Yes" (aware) and a value of 0 indicates "No" (not aware). This binary representation allows awareness (X) to be treated as a dichotomous variable - a type of categorical or nominal variable that has only two distinct, mutually exclusive values, categories, or levels suitable for categorical and probabilistic analysis. Similarly, P_i was used to represent the level of real-life adoption of cycle-syncing practices for each participant *i*. Because adoption may vary in intensity rather than being strictly present or absent, an ordinal encoding scheme was applied. Participants who reported not practicing cycle syncing were assigned a value of 0, those who practiced occasionally or inconsistently were assigned a value of 1, and those who reported regular practice were assigned a value of 2. This ordinal representation of P_i captures progressive levels of behavioral adoption while preserving the ordered nature of practice frequency. Together, the use of X for awareness and P_i for adoption provides a clear and structured representation of the key study variables, enabling meaningful clustering, comparison across groups, and subsequent statistical analysis without relying on continuous assumptions. This encoding ensured consistency

International Journal of Advance and Innovative Research
Volume 13, Issue 1 (XIII): January – March 2026
ISSN 2394 - 7780

across variables, preserved the ordinal nature of behavioral frequency, and enabled mathematical operations such as clustering, filtering, and probability estimation (Witten et al., 2016; Mitchell, 1997).

After numerical conversion, each participant i was represented using a two-dimensional feature vector:

$$X_i = (A_i, P_i) \tag{4}$$

In equation 4, A_i denotes awareness of the cycle-syncing concept (binary encoded) and P_i represents real-life adoption of cycle-syncing practices (binary or ordinal encoded). This representation allowed direct comparison across participants and supported logical filtering based on awareness adoption combinations. Filtering was subsequently applied using logical conditions derived from the encoded variables, enabling the identification of analytically relevant subgroups without modifying the total sample size. Participants were categorized based on whether they lacked awareness, possessed awareness without corresponding behavioral adoption, or actively practiced cycle syncing. In particular, filtering conditions were applied to isolate participants with low awareness, those who were aware but non-practicing, and those whose awareness translated into real-life practice. An additional subgroup capturing the awareness practice gap was identified by filtering participants who were aware of cycle syncing but reported no adoption. Logical condition-based filtering of encoded survey data is widely used in behavioral segmentation studies to examine gaps between knowledge and action (Witten et al., 2016; Mitchell, 1997).The application of this filtering framework resulted in three mutually exclusive participant groups: individuals with low awareness and no adoption of cycle-syncing practices, individuals who were aware but did not engage in consistent practice, and individuals who were both aware and actively practicing cycle syncing. The distribution of participants across these filtered groups is summarized in Table 1, which reports the corresponding sample sizes and percentages relative to the total dataset. The purpose of the filtering process was to enable meaningful subgroup analysis by separating categorical responses into analyzable numerical forms, identifying behaviorally distinct participant profiles based on cycle-syncing awareness and adoption and facilitating examination of the gap between conceptual awareness and real-life adoption. Moreover, this filtering strategy supported the application of Bernoulli-based probability modeling, descriptive statistics, and proportion-based analysis by ensuring alignment between the data structure and the statistical methodology engaged in the study (Géron, 2019; Mitchell, 1997). The filtered subsets were subsequently analyzed using frequencies, proportions, and probabilistic measures to derive interpretable insights into awareness and adoption patterns.

Table 1: Filtered Group Distribution Based on Awareness and Real-Life Adoption of Cycle-Syncing Practices

Filtered Group	Filtering Condition	Awareness (A_i)	Practice (P_i)	Number of Participants (n)	Percentage (%)
Low awareness, non-practicing	$A_i = 0 \wedge P_i = 0$	No	No	59	48.8
Aware but non-practicing	$A_i = 1 \wedge P_i = 0$	Yes	No	42	34.7
Aware and practicing	$A_i = 1 \wedge P_i \geq 1$	Yes	Yes	20	16.5
Total	—	—	—	121	100

The questionnaire used in the present study comprised 14 questions aimed at examining factors influencing the adoption of cycle-syncing practices among women. Within the dataset, one question was designated as the dependent variable, representing real-life adoption of cycle-syncing practices, while the remaining thirteen questions were treated as independent variables capturing awareness, beliefs, experiences, and menstrual health related factors. The independent variables collectively reflect respondents menstrual health awareness, perceptions of cycle syncing, personal menstrual experiences, and health-related knowledge, while the final open-ended question provided supplementary qualitative insights. This variable structure enabled a focused analysis of how menstrual health awareness and related factors influence the adoption of cycle-syncing practices.

The Bernoulli distribution is a discrete probability distribution that models a random experiment with exactly two possible outcomes, typically labeled as success and failure. A Bernoulli random variable takes the value 1 if the event of interest occurs and 0 otherwise, with a single parameter representing the probability of success. This distribution is commonly used to model binary responses such as yes/no or true/false outcomes in

 ISSN 2394 - 7780

statistical analysis and survey-based research (Ross, 2019; Devore, 2016). The Bernoulli distribution was selected as the most suitable probabilistic framework because each participant's response represents a single independent trial with mutually exclusive outcomes, and the probability of occurrence of each outcome is assumed to remain constant across participants. Moreover, the dataset does not involve repeated trials per individual or count-based events, rendering alternative distributions such as the binomial or Poisson unnecessary. The Bernoulli model also enables direct probabilistic interpretation of awareness and adoption levels, making it particularly appropriate for the present study (Ross, 2019).

The probability mass function (PMF) of a Bernoulli random variable is given by:

$$P(X = x) = p^x(1 - p)^{1-x}, x \in \{0,1\} \quad (5)$$

In equation 5, p denotes the probability of success (event occurrence), and $1 - p$ denotes the probability of failure. This formulation provides the theoretical basis for modeling binary survey responses and estimating outcome probabilities (Ross, 2019; Devore, 2016).

For a Bernoulli random variable X, the expected value and variance are defined as:

$$E(X) = p, Var(X) = p(1 - p) \quad (6)$$

In equation 6, it describes the expected value and variance of a Bernoulli random variable X, which represents a binary outcome taking the value 1 when an event occurs and 0 when it does not. The expected value E(X) is equal to p, where p denotes the probability that the event of interest occurs. This means that the average or long-run proportion of occurrences of the event is directly represented by p. The variance Var(X) is given by p (1 − p), which quantifies the degree of variability in the binary outcomes. This expression reflects the uncertainty associated with the event's occurrence: variability is highest when the probability p is moderate and decreases as p approaches either 0 or 1. When an event is almost certain or almost impossible, the variability in outcomes correspondingly diminishes. Together, the expected value and variance provide a concise statistical summary of a Bernoulli process, capturing both the average tendency of the event to occur and the dispersion around that tendency. In the context of this study, these properties support the use of Bernoulli modeling for awareness and adoption variables by formally characterizing their probabilistic behavior.

Applying these expressions to the study variables, the expected value and variance for awareness were calculated as $E(X_A) = 0.512$ and $Var(X_A) = 0.512(1 - 0.512) = 0.249$. For real-life adoption, the expected value was $E(X_P) = 0.165$, with a corresponding variance of $Var(X_P) = 0.165(1 - 0.165) = 0.138$. These measures provide insight into the central tendency and dispersion of awareness and adoption probabilities within the sample (Montgomery & Runger, 2018; Devore, 2016). The Bernoulli-based discrete distribution of the dataset is summarized in Table 2, which reports observed frequencies and corresponding probabilities for awareness and adoption outcomes across the sample. The probabilistic representation indicates that while awareness of cycle syncing is present among a moderate proportion of participants, real-life adoption remains substantially lower. This distributional pattern highlights a pronounced awareness practice gap, reinforcing the need for further statistical testing to examine the relationship between awareness and behavioral adoption, which is central to the study's analytical objectives.

Table 2: Bernoulli-Based Discrete Data Distribution (n = 121)

Variable	Outcome (x)	Frequency (n)	Probability $P(X = x)$	Distribution Type
Awareness of cycle syncing (X_A)	1 (Yes) 0 (No)	62 59	$p_A = 0.512$ $1 - p_A = 0.488$	Bernoulli
Real-life adoption (X_P)	1 (Yes) 0 (No)	20 101	$p_P = 0.165$ $1 - p_P = 0.835$	Bernoulli

The present study engaged a non-probability convenience sampling technique for primary data collection, followed by a cluster-based analytical sampling framework for subgroup-level analysis. Female participants who voluntarily responded to the structured online questionnaire distributed via Google Forms constituted the initial sample. Rather than drawing separate samples for different subgroups, the study adopted a post-survey analytical sampling approach using clustering techniques. This approach enabled segmentation of participants based on similarities in key study variables without altering the original sample size or introducing additional sampling bias, which is particularly appropriate for behavioral and health-related survey research.

ISSN 2394 - 7780

Analytical sampling within the study was operationalized using the K-means clustering algorithm, an unsupervised machine learning technique that partitions observations into a predefined number of clusters by minimizing within-cluster variance and maximizing similarity among data points within each cluster.

K-means clustering is widely used to identify underlying patterns and groupings in survey data based on shared response characteristics (MacQueen, 1967; Hastie, Tibshirani, & Friedman, 2009).

which grouped participants based on two core variables central to the study objectives: awareness of the cycle-syncing concept and real-life Each participant was represented as a two-dimensional feature vector, as defined earlier in equation 4, where awareness and adoption variables were numerically encoded to enable mathematical operations. This encoding ensured consistency across variables, preserved the behavioral meaning of the responses, and supported analytical procedures such as clustering, filtering, and probability estimation (Witten et al., 2016; Mitchell, 1997). Following numerical representation, logical condition-based filtering was applied to identify analytically relevant subgroups without altering the total sample size. Participants were categorized based on combinations of awareness and adoption status, including individuals with low awareness, those who were aware but did not practice cycle syncing, and those whose awareness translated into real-life adoption.

An additional subgroup representing the awareness practice gap was identified by isolating participants who were aware of cycle syncing but reported no behavioral adoption. Such logical filtering of encoded survey data is widely applied in behavioral segmentation research to examine discrepancies between knowledge and action (Witten et al., 2016; Mitchell, 1997).

Cluster assignment was performed using the minimum Euclidean distance between each participant's feature vector and the cluster centroids, defined using Equation 7:

$$d(X_i, C_j) = \sqrt{(A_i - \mu_{Aj})^2 + (P_i - \mu_{Pj})^2} \qquad (7)$$

In equation 7, C_j denotes the centroid of cluster j, and μ_{Aj} and μ_{Pj} represent the mean awareness and adoption scores of that cluster. The clustering algorithm iteratively updated centroid positions with the objective of minimizing within-cluster variance, expressed using Equation 8:

$$J = \sum_{j=1}^{k} \sum_{i \in C_j} \| X_i - \mu_j \|^2 \qquad (8)$$

In equation 8, $k = 3$ clusters were specified based on theoretical relevance to awareness–adoption progression and interpretability of behavioral segmentation. The formulation of K-means clustering, Euclidean distance–based assignment, and variance minimization is grounded in classical unsupervised learning theory and has been extensively applied in behavioral segmentation research (MacQueen, 1967; Hastie et al., 2009). This clustering-based analytical sampling framework preserved the entire sample size n = 121, enabled data-driven subgroup formation, and facilitated comparison between conceptual awareness and real-life behavioral adoption without introducing additional sampling bias. Clustering was therefore applied strictly as an analytical sampling mechanism rather than a data collection strategy.

Table 3: Cluster-Wise Distribution of Awareness and Real-Life Adoption of Cycle-Syncing Practices (n = 121)

Cluster	Sampling Condition	Awareness of Cycle Syncing (A_i)	Real-Life Adoption (P_i)	Sample Size (n)	Percentage (%)	Sampling Interpretation
Cluster 0	$A_i = 0 \land P_i = 0$	Not aware	Not practicing	59	48.8	Participants with low awareness and no adoption
Cluster 1	$A_i = 1 \land P_i = 0$	Aware	Not practicing / inconsistent	42	34.7	Awareness present but behavior not adopted
Cluster 2	$A_i = 1 \land P_i \geq 1$	Aware	Actively practicing	20	16.5	Awareness translated into real-life practice
Total	—	—	—	121	100	—

Table 3 presents the cluster-wise distribution of participants based on awareness of the cycle-syncing concept and its real-life adoption among the study sample (n = 121). The clustering is performed using binary classifications of awareness (A_i) and practice/adoption (P_i), where a value of 1 indicates presence and 0

ISSN 2394 - 7780

indicates absence. This approach enables segmentation of participants into behaviorally and cognitively distinct groups, facilitating a structured interpretation of the awareness–adoption relationship. Cluster 0 ($A_i = 0$, $P_i = 0$) represents participants who are neither aware of cycle syncing nor practicing it.

This group constitutes the largest proportion of the sample (n = 59; 48.8%). The cluster reflects individuals with low exposure and no behavioral engagement, serving as a baseline group indicating absence of both knowledge and adoption. Cluster 1 ($A_i = 1$, $P_i = 0$) includes participants who are aware of the cycle-syncing concept but do not practice it consistently or at all. This cluster accounts for 42 participants (34.7%). The presence of this group highlights a knowledge–behavior gap, where awareness does not automatically translate into real-life adoption. This cluster is particularly important for understanding barriers to behavioral change despite informational exposure. Cluster 2 ($A_i = 1$, $P_i = 1$) consists of participants who are both aware of cycle syncing and actively practicing it. This is the smallest cluster (n = 20, 16.5%) and represents successful translation of awareness into real-life behavioral adoption. This group provides empirical support for the study's alternative hypothesis, indicating that awareness can lead to adoption, albeit for a limited subset of participants. Overall, the clustering approach demonstrates that while awareness of cycle syncing exists among a substantial proportion of participants, actual adoption remains comparatively low. The dominance of Cluster 1 suggests that awareness alone is insufficient for consistent behavioral change, underscoring the need to examine contextual, motivational, and practical factors influencing adoption. Thus, Table 3 offers a structured empirical foundation for subsequent hypothesis testing and statistical analysis of the awareness adoption relationship.

Figure 1 illustrates the K-means clustering workflow applied in this study to segment participants based on cycle-syncing awareness and real-life adoption. Each participant is represented as a two-dimensional feature vector, iteratively assigned to the nearest cluster centroid using Euclidean distance, with centroids updated to minimize within-cluster variance. This architecture follows the original K-means formulation proposed by MacQueen (1967) and its standard optimization-based interpretation in unsupervised learning (Hastie, Tibshirani, & Friedman, 2009).

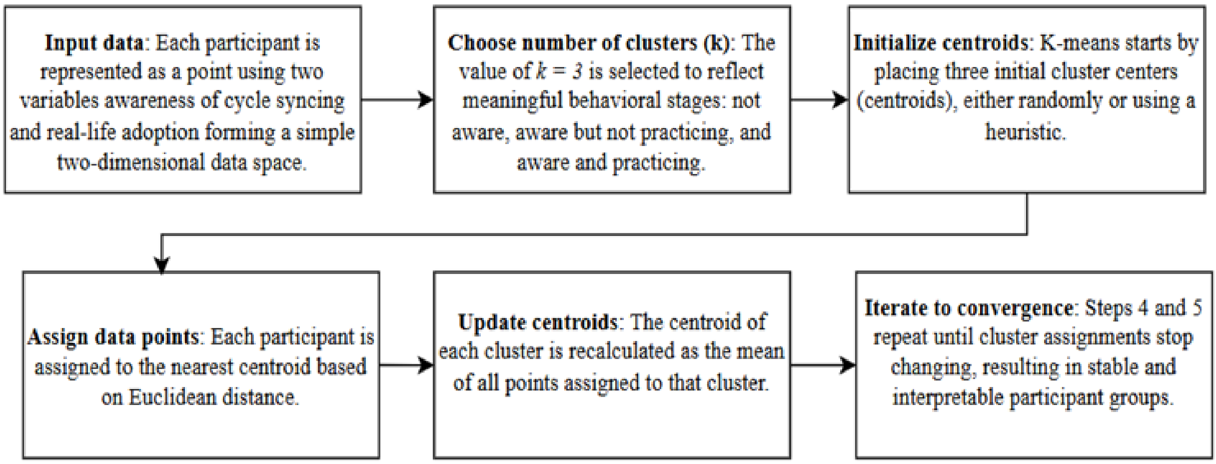

Figure 1: K-means Clustering Architecture for Awareness–Adoption Segmentation

The statistical analysis was conducted to evaluate the difference between awareness of the cycle-syncing concept and its real-life adoption among women using proportion-based inference. As both variables were binary in nature (Yes/No) and previously modelled as Bernoulli random variables, a two-proportion test based on the chi-square (χ^2) method was employed to compare the observed proportions within the same study population. Proportion-based testing is particularly suitable in this context because it allows direct comparison of probabilities derived from discrete binary outcomes commonly observed in behavioral and health-related survey research (Ross, 2019; Montgomery & Runger, 2018).

In equation 9, Let p_A denote the proportion of participants aware of the cycle-syncing concept and p_P denote the proportion of participants who actively practiced cycle syncing in real life as earlier discussed. The analysis focused on assessing whether the proportion of awareness exceeded the proportion of adoption, in alignment with the analytical framework and hypotheses established earlier. The test was performed using observed counts from the filtered dataset of 121 participants. Of these, 62 participants reported awareness of cycle syncing, while 20 participants reported active adoption. Accordingly, the estimated proportions were computed as:

$$p_A = \frac{62}{121} = 0.512, p_P = \frac{20}{121} = 0.165 \quad (9)$$

ISSN 2394 - 7780

To formally test the difference between these proportions, a two-proportion chi-square test was implemented using the prop.test() function in R with a one-sided alternative hypothesis. This test is grounded in the Pearson chi-square statistic, which measures the extent to which observed frequencies deviate from expected frequencies under the null hypothesis of equal proportions. The chi-square test statistic is defined as:

$$\chi^2 = \sum_{i=1}^{k} \frac{(O_i - E_i)^2}{E_i} \qquad (10)$$

In equation 10, O_i represents the observed frequencies and E_i represents the expected frequencies under the null assumption that awareness and adoption occur at the same rate. The chi-square framework is particularly appropriate for large-sample proportion testing, as it provides a reliable approximation for assessing differences between categorical distributions (Montgomery & Runger, 2018; Agresti, 2019).

Expected frequencies in the two-proportion test are computed using the pooled proportion, given by:

$$\hat{p} = \frac{x_1 + x_2}{n_1 + n_2} \qquad \text{Equation (11)}$$

In equation 11, x_1 and x_2 denote the number of successes in each group, while n_1 and n_2 represent the corresponding sample sizes. The pooled estimator is a standard component of two-proportion testing and ensures unbiased estimation of expected frequencies under the null hypothesis of equal proportions (Casella & Berger, 2002; Montgomery & Runger, 2018).

The statistical test was executed using the following R command:

```
prop.test(
x = c(62, 20),
n = c(121, 121),
alternative = "greater"
)
```

The prop.test() function computes the chi-square test statistic and the associated p-value using a normal approximation with continuity correction, which is recommended for proportion comparisons involving moderate to large sample sizes (Montgomery & Runger, 2018; Agresti, 2019). A one-sided alternative hypothesis was specified to test whether awareness of cycle syncing was significantly greater than its real-life adoption. The chi-square method was particularly well-suited for this study because it accommodates binary categorical data, relies on frequency-based comparisons, and aligns naturally with the Bernoulli modeling framework established earlier. Moreover, it enables direct testing of the awareness–adoption gap without imposing assumptions of normality on individual responses. This approach ensured methodological consistency, supported probabilistic interpretation of observed proportions, and provided a robust inferential framework for evaluating the relationship between awareness and behavioral adoption of cycle-syncing practices within the study population.

4. RESULT & DISCUSSION

The hypothesis of our research paper is to examine whether awareness of the cycle-syncing concept influences the real-life adoption of cycle-syncing practices among women. The study aims to understand if simply being informed about cycle-syncing translates into meaningful behavioral changes in daily lifestyle choices related to menstrual health. The null hypothesis (H_0) states that awareness of the cycle-syncing concept has no effect on its practical adoption. In this view, women who are aware of cycle-syncing are no more likely to adopt cycle-syncing behaviors than those who are unaware, indicating that knowledge alone does not influence practice. In contrast, the alternative hypothesis (H_1) proposes that awareness of the cycle-syncing concept positively influences its adoption in real life. This hypothesis assumes that women who are informed about cycle-syncing are more likely to incorporate cycle-aligned behaviors into their routines, reflecting a meaningful link between menstrual health awareness and actionable lifestyle practices.

The results of the two-proportion statistical analysis revealed a substantial difference between awareness of the cycle-syncing concept and its real-life adoption among the study participants. The test produced a chi-square statistic of $\chi^2 = 31.006$ with 1 degree of freedom, indicating a strong deviation from the assumption of equal proportions. The associated p-value was 1.286×10^{-8}, which is considerably lower than the conventional significance threshold of 0.05. The observed sample proportions showed that 51.2% of participants reported being aware of the cycle-syncing concept, whereas only 16.5% reported actively practicing cycle syncing in real life. This large discrepancy between awareness and adoption was further supported by the 95% confidence interval for the difference in proportions, which ranged from 0.246 to 1.000. As the confidence interval did not include zero, the difference between awareness and adoption proportions can be considered statistically significant, providing additional evidence against the null assumption of equal proportions (Montgomery & Runger, 2018; Agresti, 2019). Table 4 summarizes the two-proportion chi-square test comparing awareness of the cycle-syncing concept ($prop_1$) and its real-life adoption ($prop_2$).

Table 4: Two-Proportion Chi-Square Test Comparing Awareness and Real-Life Adoption of Cycle Syncing

Measure	Awareness of Cycle Syncing ($prop_1$)	Real-Life Adoption of Cycle Syncing ($prop_2$)
Number of successes (x)	62	20
Sample size (n)	121	121
Sample proportion	0.512	0.165
Percentage (%)	51.2%	16.5%

These findings indicate that awareness of cycle syncing is significantly more prevalent than its real-life implementation among women in the study population. Although awareness is present in a substantial proportion of participants, this knowledge does not consistently translate into behavioral adoption, highlighting a pronounced awareness practice gap. These results highlight the importance of moving beyond awareness-based education toward behavior-oriented menstrual health literacy. Providing women with structured, practical strategies for managing daily activities in alignment with menstrual cycle phases is essential for encouraging regular adoption of cycle-syncing practices. Educational interventions and digital health tools should therefore emphasize not only what cycle syncing is, but also how it can be realistically incorporated into daily life. Strengthening applied menstrual literacy in this manner may reduce the observed awareness–practice gap and promote sustained engagement with cycle-aligned lifestyle practices.

The statistical analysis revealed a clear and statistically significant difference between awareness of the cycle-syncing concept and its real-life adoption among the study participants. The magnitude of the chi-square test statistic, together with the associated p-value well below the conventional significance threshold, provides strong evidence against the assumption of equal proportions. This indicates that the observed disparity between awareness and adoption is unlikely to have occurred by random variation alone. Furthermore, the confidence interval for the difference in proportions did not include zero and demonstrated a consistent positive direction, reinforcing the stability and robustness of the observed effect. Collectively, these findings provide quantitative support for the presence of a pronounced awareness practice gap, highlighting that while knowledge of cycle syncing is relatively widespread, its translation into consistent real-world behavior remains limited.

Overall, the results establish that while awareness of the cycle-syncing concept is relatively widespread, real-life adoption remains limited within the sampled population. Based on the results of the two-proportion test, the null hypothesis (H_0) was rejected, and the alternative hypothesis (H_1) was accepted. The findings confirm that awareness of the cycle-syncing concept is significantly higher than its real-life adoption among women. While awareness appears to increase the likelihood of adoption, a substantial gap remains between knowledge and consistent behavioral implementation. Beyond confirming the statistical significance of the awareness–adoption gap, these findings underscore the complexity of translating menstrual health knowledge into sustained behavioral change. The acceptance of the alternative hypothesis indicates that awareness is a necessary but not sufficient condition for real-life adoption of cycle-syncing practices. While informed individuals demonstrate a higher propensity toward adoption, the persistence of a large aware yet non-practicing subgroup suggests the presence of intervening factors that moderate this relationship. Practical constraints, limited phase-specific understanding, and lack of structured guidance may hinder consistent implementation. Thus, awareness alone does not ensure adoption unless supported by applied, context-sensitive strategies. These findings indicate that the awareness–adoption relationship is probabilistic, emphasizing menstrual health literacy as a process of knowledge, interpretation, and action.

5. CONCLUSION

The present study examined women's menstrual health literacy with a specific focus on awareness and real-life adoption of the cycle syncing concept. Using a descriptive, survey-based empirical approach, the study assessed whether conceptual knowledge regarding menstrual cycle phases and cycle syncing translated into consistent lifestyle practices and attentiveness toward menstrual irregularities. The findings demonstrated that while awareness levels were notably high, practical implementation remained limited among a substantial proportion of women. The results revealed that a majority of participants were aware of the four phases of the menstrual cycle and the concept of cycle syncing. However, nearly half of the respondents did not actively follow cycle syncing practices in their daily lives, indicating a clear gap between knowledge and behavior. In addition, a significant proportion of women reported experiencing menstrual irregularities such as late or missed periods, even though many acknowledged the impact of these irregularities on their physical health, emotional well-being, and daily functioning. Awareness related to hormonal health conditions, including PCOD, was also high, yet proactive lifestyle alignment and consistent cycle-based self-management were not uniformly practiced.

These findings highlight that menstrual health literacy alone is insufficient to ensure behavioral adoption of cycle-aware practices. Socio-cultural factors, perceived barriers, limited guidance, and lack of structured support may contribute to this disconnect between awareness and action. The study underscores the need for targeted educational initiatives and accessible digital health interventions that not only raise awareness but also facilitate practical implementation of cycle syncing in everyday life. Overall, this research contributes empirical evidence to the emerging discourse on cycle syncing by demonstrating the disparity between awareness and real-life practice. The findings emphasize the importance of moving beyond knowledge dissemination toward behavior-focused strategies to improve menstrual health outcomes and support women in managing menstrual irregularities more effectively. To position the contribution of the present study within the existing body of menstrual health and cycle-syncing research, a comparative analysis was conducted against prior studies focusing on menstrual health literacy, digital cycle-tracking technologies, and behavioral engagement. While earlier research has extensively examined awareness, knowledge improvement, and app-based engagement, limited studies have quantitatively assessed the translation of awareness into real-life cycle-syncing practices using probabilistic and statistical modeling. Table 5 summarizes this comparison explaining my work compared with previous research.

Table 5: Comparison of the Present Study with Existing Research

Author(s)	Model / Methodology	Focus of Their Work	Key Findings / Results
Cunningham et al. (2024)	Survey-based evaluation, descriptive statistics	Assessed the effectiveness of the Flo app in improving menstrual health literacy and menstrual knowledge	Found that use of the Flo app was associated with improved menstrual health literacy and increased awareness of menstrual cycle concepts among users
Goddard et al. (2025)	Longitudinal survey analysis	Examined menstrual knowledge and health outcomes associated with cycle-tracking app usage	Reported sustained improvements in menstrual knowledge and selected health outcomes over time among users of cycle-tracking applications
Schantz et al. (2021)	Comprehensive literature review	Reviewed menstrual cycle tracking applications and their epidemiological potential	Highlighted the growing role of menstrual tracking technologies in data-driven menstrual health research and population-level insights
Baird (2022); Hennegan et al. (2019)	Educational intervention analysis	Investigated improvements in menstrual health literacy through life-skills education	Demonstrated that structured educational interventions significantly enhance menstrual health knowledge and awareness
Stujenske (2023); Epstein et al. (2017)	Survey analysis of menstrual cycle technology use	Analyzed usage patterns and engagement with menstrual cycle technologies	Identified common patterns of engagement, self-tracking behaviors, and user interaction with menstrual health technologies
Present Study	Bernoulli distribution modeling, clustering, filtering, two-proportion chi-square test	Examined the relationship between awareness of cycle syncing and its real-life adoption	Empirically identified a significant gap between awareness and real-life adoption of cycle-syncing practices, demonstrating that increased awareness does not necessarily translate into behavioral implementation

In table 5, the present study contributes to the menstrual health literature by empirically examining the relationship between awareness of the cycle-syncing concept and its real-life adoption among women. Consistent with prior research emphasizing the growing availability of menstrual health information and digital tracking tools (Cunningham et al., 2024; Goddard et al., 2025; Schantz et al., 2021), the findings confirm that awareness of cycle syncing is relatively widespread within the study population. However, the results also reveal that this awareness does not consistently translate into behavioral adoption, highlighting a substantial awareness practice gap. The earlier studies that primarily focused on improving menstrual health literacy (Baird, 2022; Hennegan et al., 2019), app-based engagement (Stujenske, 2023; Epstein et al., 2017), or conceptual discussions of cycle syncing (Francois, 2024), but the present study employed a probabilistic and statistical framework to quantify this gap. By modeling awareness and adoption as Bernoulli random variables and applying a two-proportion chi-square test, the study demonstrated that the proportion of women aware of cycle syncing was significantly higher than the proportion actively practicing it in real life.

This finding reinforces existing evidence that knowledge acquisition alone is insufficient to ensure sustained behavioral change in menstrual health contexts (Schantz et al., 2021; Chen et al., 2023). The results underscore the need for interventions that move beyond information dissemination toward practical guidance, contextual support, and behavior-focused strategies that facilitate real-world implementation of cycle-syncing practices. Overall, the study strengthens the existing literature by providing statistically validated evidence of the awareness practice disconnect in cycle syncing. By integrating survey data with probability-based modeling and inferential analysis, the present work offers a methodologically rigorous contribution that complements prior descriptive and conceptual studies and provides a foundation for future research aimed at improving behavioral adoption of menstrual cycle-aligned health practices.

Future Work & Recommendation

The future research should expand on the present findings by exploring the cycle-syncing concept in greater depth, with particular emphasis on its applicability and effectiveness among women with menstrual and hormonal disorders such as polycystic ovarian disease (PCOD). While the current study examined awareness and general adoption patterns, subsequent studies could investigate how phase-specific lifestyle adjustments in nutrition, physical activity, stress management, and self-care influence symptom severity, cycle regularity, and overall well-being in women affected by PCOD. Applying longitudinal designs, clinically validated symptom measures, and collaboration with healthcare professionals would allow for a more comprehensive evaluation of cycle syncing as a supportive, non-pharmacological approach to menstrual health management.

Acknowledgement

The authors would like to express their sincere gratitude to Sheth L. U. Jhaveri College, Mumbai, for providing the academic environment, institutional support, and resources necessary for the successful completion of this research.

Ethical Statement

This study involved the collection of data from human participants through an anonymous, voluntary online survey. The research was conducted in accordance with ethical standards for social science research. Participation was entirely voluntary, and informed consent was obtained from all participants prior to data collection. No personal identifiers were collected, and participant anonymity and confidentiality were maintained throughout the study. The data were used solely for academic and research purposes.

Conflicts of Interest

The authors declare that they have no conflicts of interest to this work.

Data Availability Statement

(a) Data sharing is not applicable to this article as no new data were created or analyzed in this study.

REFERENCES

Cunningham, A. C., Ballard, A. M., Zhaunova, L., & Prentice, C. (2024). Efficacy of the Flo app in improving health literacy and menstrual knowledge. JMIR mHealth and uHealth, 12, e54124. https://mhealth.jmir.org/2024/1/e54124

Maharashtra State Bureau of Textbook Production and Curriculum Research. (2022). *Class 12 Biology textbook* [PDF]. Maharashtra State Board of Secondary & Higher Secondary Education. https://static.collegedekho.com/media/uploads/2022/09/21/class-12-biology-textbook-pdf-1.pdf

Goddard, F. G. B., Prentice, C., Cunningham, A. C., Ballard, A. M., & Zhaunova, L. (2025). Impact of the Flo cycle tracking app on menstrual knowledge and health in low-income and middle-income countries: A longitudinal study. BMJ Public Health, 3(2). https://doi.org/10.1136/bmjph-2025-002822

Baird, S. (2022). Improving menstrual health literacy through life-skills education. Frontiers in Global Women's Health, 3, 838961. https://www.frontiersin.org/articles/10.3389/fgwh.2022.838961/full

Stujenske, T. M. (2023). Survey analysis of menstrual cycle technology use. Medicina, 59(9), 1509. https://www.mdpi.com/1648-9144/59/9/1509

Gopalan, S. S., Mishra, A., Ramaswami, A., & Subramanian, S. V. (2024). Impact of menstrual symptoms, experiences, and perceptions on women's health and well-being. Preventive Medicine, 176, 107792. https://www.sciencedirect.com/science/article/pii/S009174352400135X

Francois, M. (2024). Cycle syncing: Optimizing women's quality of life (Master's thesis, Augsburg University, Minnesota, USA). https://idun.augsburg.edu/etd/1607/

Saini, R. K., & Dutta, U. (2025). Rhythms of the menstrual cycle in therapy rooms: Integrating hormonal awareness into therapeutic practice. Annals of Indian Psychiatry. https://journals.lww.com/aips/fulltext/9900/rhythms_of_the_menstrual_cycle_in_therapy_rooms_.130.aspx

Pfender, E., Wanzer, C., Mikkers, L., & Bleakley, A. (2025). Sync or swim: Navigating the tides of menstrual cycle messaging on TikTok. Perspectives on Sexual and Reproductive Health. https://doi.org/10.1111/psrh.70004

Signos. (2024). Cycle syncing for metabolic health: How to align diet and activity to your menstrual cycle. https://www.signos.com/blog/cycle-syncing-for-metabolic-health

Itriyeva, K. (2022). The normal menstrual cycle. Current Problems in Pediatric and Adolescent Health Care, 52(5), 101183. https://doi.org/10.1016/j.cppeds.2022.101183

Kapitula, M., Hendrickson, A., & Lalli, S. (2023). Assessing women's understanding of menstruation: Knowledge gaps and educational needs (Student research poster, Misericordia University). https://digitalcommons.misericordia.edu/research_posters2023/16/

Schantz, J. S., Fernandez, C. S. P., & Jukic, A. M. Z. (2021). Menstrual cycle tracking applications and the potential for epidemiological research: A comprehensive review of the literature. Current Epidemiology Reports, 8, 9–19. https://doi.org/10.1007/s40471-020-00260-3

Chen, K. H., Lin, M. L., & Huang, Y. C. (2023). Barriers and facilitators to sustained engagement with mobile menstrual health applications: A mixed-methods study. JMIR Formative Research, 7, e46821. https://mhealth.jmir.org/2019/4/e11223/citations

Hennegan, J., Shannon, A. K., Rubli, J., Schwab, K. J., & Melendez-Torres, G. J. (2019). Women's and girls' experiences of menstruation in low- and middle-income countries: A systematic review and qualitative metasynthesis. PLOS Medicine, 16(5), e1002803. https://doi.org/10.1371/journal.pmed.1002803

Epstein, D. A., Ping, A., Fogarty, J., Munson, S. A., & Kientz, J. A. (2017). A lived informatics model of personal informatics. Proceedings of the ACM on Human-Computer Interaction, 1(CSCW), 1–19. https://dl.acm.org/doi/10.1145/2750858.2804250

Ross, S. M. (2019). Introduction to probability models (12th ed.). https://shop.elsevier.com/books/introduction-to-probability-models/ross/978-0-12-814346-9

MacQueen, J. (1967). Some methods for classification and analysis of multivariate observations. In Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability (Vol. 1, pp. 281–297). University of California Press. https://projecteuclid.org/ebooks/berkeley-symposium-on-mathematical-statistics-and-probability/Proceedings-of-the-Fifth-Berkeley-Symposium-on-Mathematical-Statistics-and/chapter/Some-methods-for-classification-and-analysis-of-multivariate-observations/bsmsp/1200512992

Hastie, T., Tibshirani, R., & Friedman, J. (2009). The elements of statistical learning: Data mining, inference, and prediction (2nd ed.). Springer. https://hastie.su.domains/ElemStatLearn/

Géron, A. (2019). Hands-on machine learning with Scikit-Learn, Keras, and TensorFlow (2nd ed.). O'Reilly Media. https://www.oreilly.com/library/view/hands-on-machine-learning/9781492032632/

Witten, I. H., Frank, E., Hall, M. A., & Pal, C. J. (2016). Data mining: Practical machine learning tools and techniques (4th ed.). Morgan Kaufmann. https://shop.elsevier.com/books/data-mining/witten/978-0-12-804291-5

Mitchell, T. M. (1997). Machine learning. McGraw-Hill. https://cdn.chools.in/DIG_LIB/E-Book/M1-Machine-Learning-Tom-Mitchell_.pdf

Casella, G., & Berger, R. L. (2002). Statistical inference (2nd ed.). Duxbury Press. https://books.google.co.in/books?id=0x_vAAAAMAAJ&redir_esc=y

Montgomery, D. C., & Runger, G. C. (2018). Applied statistics and probability for engineers (7th ed.). Wiley. https://www.wiley.com/en-us/Applied+Statistics+and+Probability+for+Engineers%2C+7th+Edition-p-9781119400363

Devore, J. L. (2016). Probability and statistics for engineering and the sciences (9th ed.). Cengage Learning. https://cdn.prexams.com/9441/Devore%20Solutions%20Ed9.pdf

Agresti, A. (2019). An introduction to categorical data analysis (3rd ed.). Wiley.

https://www.wiley.com/en-us/An+Introduction+to+Categorical+Data+Analysis%2C+3rd+Edition-p-9781119405283