Engagement with Mental Health Screening on Mobile Devices: Results from an Antenatal Feasibility Study

Kevin Doherty
School of Computer Science and Statistics, Trinity College Dublin, Ireland

José Marcano-Belisario
School of Public Health, Imperial College London, London, United Kingdom

Martin Cohn
School of Public Health, Imperial College London, London, United Kingdom

Nikolaos Mastellos
School of Public Health, Imperial College London, London, United Kingdom

Cecily Morrison
Human Experience & Design, Microsoft Research Cambridge, Cambridge, United Kingdom

Josip Car
Global eHealth Unit, Department of Primary Care and Public Health, Imperial College London, London, United Kingdom

Gavin Doherty
School of Computer Science and Statistics, Trinity College Dublin, Ireland
Gavin.Doherty@scss.tcd.ie

ABSTRACT

Perinatal depression (PND) affects up to 15% of women within the United Kingdom and has a lasting impact on a woman's quality of life, birth outcomes and her child's development. Suicide is the leading cause of maternal mortality. However, it is estimated that at least 50% of PND cases go undiagnosed. This paper presents the results of the first feasibility study to examine the potential of mobile devices to engage women in antenatal mental health screening. Using a mobile application, 254 women attending 14 National Health Service midwifery clinics provided 2,280 momentary and retrospective reports of their wellbeing over a 9-month period. Women spoke positively of the experience, installing and engaging with this technology regardless of age, education, wellbeing, number of children, marital or employment status, or past diagnosis of depression. 39 women reported a risk of depression, self-harm or suicide; two-thirds of whom were not identified by screening in-clinic.

ACM ISBN 978-1-4503-5970-2/19/05.
https://doi.org/10.1145/3290605.3300416

CCS CONCEPTS

• **Human-centered computing → Empirical studies in HCI**; *Empirical studies in interaction design*; *Mobile devices.*

KEYWORDS

Pregnancy, Public Health Screening, Mental Health, Psychological Wellbeing, Self-Report, EMA, Mobile Devices

ACM Reference Format:
Kevin Doherty, José Marcano-Belisario, Martin Cohn, Nikolaos Mastellos, Cecily Morrison, Josip Car, and Gavin Doherty. 2019. Engagement with Mental Health Screening on Mobile Devices: Results from an Antenatal Feasibility Study. In *CHI Conference on Human Factors in Computing Systems Proceedings (CHI 2019), May 4–9, 2019, Glasgow, Scotland UK*. ACM, New York, NY, USA, 15 pages. https://doi.org/10.1145/3290605.3300416

1 INTRODUCTION

Perinatal depression (PND) affects up to 15% of women during pregnancy or within one year of giving birth in the United Kingdom (UK) [5]. Pregnant women suffering from depression are more likely to engage in unhealthy practices including poor diet, substance abuse and failure to enrol in prenatal care, and are at increased risk of self-harm, suicide and postnatal depression [9, 32, 48, 80]. Suicide is the leading cause of maternal mortality within the UK [41]. Antenatal depression can also affect fetal development and has been identified as an independent risk factor for children's development through adolescence [13, 40, 62, 80].

Treating depression during pregnancy can reduce the likelihood of developing postnatal depression, prevent more severe forms of the condition, reduce its intergenerational impact and improve women's overall health status [41, 80]. Treatment and support needs to be made available to those who need it, but in order to do so, effective programs of assessment, particularly for those at risk of distress, need to be in place. In the context of the UK's National Health Service (NHS), mental health screening in pregnancy is currently carried out verbally and using paper based questionnaires completed in waiting rooms. 96% of UK midwives report asking women about their mental wellbeing at their first appointment [12]. However, only one in ten women recall being asked (ibid.). Less than a fifth of women report being completely honest with professionals and a third report never sharing that they felt unwell during pregnancy [12]. It is estimated that at least 50% of PND cases go undiagnosed [67, 83]. Women report that they refrain from initiating discussions with professionals about their mental health, or provide inaccurate responses to screening scales, due to discomfort, stigma and uncertainty around the expected emotional experience of pregnancy [16, 39].

The challenges of mental healthcare provision have been described as "so long standing, so vast, and so unresponsive" that they require a new approach entailing levers of change on a population scale [6]. Mobile devices have the potential to facilitate the self-report and remote screening of mood and depression throughout the antenatal period, extending care to under-served and at-risk populations, enabling timely assessment and intervention, supporting honest disclosure, and fostering trust between women and midwives. However, we know little about the feasibility of deploying these technologies for public health screening in practice.

This paper details the results of the first longitudinal deployment of a mobile application for the self-report of psychological wellbeing in antenatal clinical practice; an interdisciplinary undertaking involving HCI and public health researchers, pregnant women and a variety of health professionals including midwives. This feasibility study examines the capacity of mobile technologies to engage pregnant women in the self-report of wellbeing and depression in daily life, facilitating treatment and support for those in distress.

2 RELATED WORK

Mobile Technologies for Perinatal Wellbeing

Mobile devices have long been thought to possess the potential to transform research, clinical practice and wellbeing at a population scale [6, 24, 46, 76, 91]. Millions of women have installed thousands of mobile applications in the hope of supporting a healthy start to life. The majority of these technologies have been designed to communicate health-related information to parents [86]. HCI researchers have developed prototype applications for Dutch (Babywijzer), Pakistani (Baby+), and Vietnamese Australian (We-HELP) populations [70, 78, 92], deployed SMS-based systems for personalised health information communication in Kenya and Pakistan [8, 63] and conducted qualitative analyses of pregnant women's motivations for information sharing and support seeking online [29, 43, 65]. Peyton et al. propose a 'pregnancy ecology,' comprising physical, emotional, informational and social supports, to support the design of physical health interventions [64], which Prabhakar et al. extend to include support needs, sources, and interventions within an Evolving Ecology of Support [66].

Researchers have also explored the design of prototype technologies for health data tracking in pregnancy, including nutrition, hydration, activity, weight and mood (Bloom) [90], nausea and vomiting (Dot-it) [47], and physiological data (Nuwa) [27]. Other perinatal research has examined women's motivations with respect to menstrual tracking applications [26] and the disclosure of pregnancy loss on social networks [3], the prediction of postnatal depression from survey data [59], the use of twitter to track developmental milestones in young children [82], and the design of applications for monitoring the health of preterm babies (Estrellita) [30], sharing infant activity data with friends and family (MammiBelli) [36], and to support breast-milk donation (Milk Matters) [89].

Design research has facilitated knowledge of women's and other stigmatised groups' needs with respect to a variety of prototype systems [49]. However, less HCI research has attended to the real-world use of technology in the perinatal context, to the role midwives and other health professionals play in pregnancy, to the integration of personal devices within a clinical and public health context, to the subjective experience of pregnancy or to the significant implications of maternal mental health.

The Practice of Self-Report in Daily Life

Much of the potential of mobile devices to support wellbeing stems from their capacity to facilitate an understanding of subjective experience in daily life. This has been described as a form of systematic phenomenology [31], capturing "life as it is lived, moment to moment, hour to hour, day to day" [76] and permitting the "study of the stream of thought or behaviour" [34]. The methodology of self-report in daily life is most often referred to as Ecological Momentary Assessment (EMA) or the Experience Sampling Method (ESM) [73, 76]. HCI researchers have designed EMA technologies for Parkinson's disease [88], post-traumatic stress disorder [45], asthma [15], pain [1], stress and sleepiness [61], bipolar disorder [53], depression [52], anxiety [84] and more.

In fact, much of what we know about mental health and wellbeing is gathered through self-report. Health professionals' knowledge of patients' wellbeing is typically retrospective in nature, drawn from clinical interviews and validated screening questionnaires. However, several decades of cognitive psychology and behavioural economics research has revealed striking differences between reports of wellbeing made in the moment and looking back over time [14, 25, 51, 76, 79, 81]. Demonstrating the real-world implications of such findings, one study (n=108) found that the variability, rather than intensity, of depressive affect predicted suicidal ideation and suicide attempts among college students [93], and another (n=36) found that momentary measures of affect predicted self-harm behaviours [4].

The assessment of wellbeing is therefore subject to questions with different time-frames [22]. While retrospective or trait measures can be 'too cold,' "too slow and sluggish to change," momentary measures can also prove 'too hot,' "too volatile and overly sensitive to extraneous variables" [19]. Designers of EMA sampling protocols must therefore attend to the subjective experience of self-report, the temporal distinction of reporting mechanisms, the collection of valid data, and the burden of reporting. There currently exists "a paucity of studies comparing various time-based designs" and "no general conventions," a fact which Santangelo et al. write is unsurprising, "as the temporal dynamics of emotional and cognitive processes are largely unknown" [72].

Large-Scale and Longitudinal Patient Engagement

The feasibility of self-report technologies hinges in large part upon the engagement of users [23, 85]. Studies often report a swift decline in reporting practice following "an initial burst of interest" [10, 17, 18]. Designing for engagement with respect to self-report technologies typically entails reducing the burden on users by implementing a simple and efficient reporting process while at the same time providing intrinsic and extrinsic incentives to users, and attending to the challenges of reactivity, habituation, validity and data completeness [18, 25, 42, 44, 55, 60]. The merit of a variety of incentives, including monetary remuneration, booster telephone calls, reminder emails, data visualisation and feedback, remains an active topic of research [68].

Mental health screening in pregnancy is further complicated by the need to support disclosure while overcoming a pervasive mental health-related stigma [28, 35, 39]. In previous design research, the suggestion that pregnant women should "actively track and record their activities," was "met with incredulity, laughter and sometimes derision" [64]. Depression during pregnancy is marked by an unwillingness to seek help at what parents believe should be a happy time [58]. Concerns therefore often go unreported, despite heightened attention to wellbeing.

Mobile devices have the potential to overcome the public health challenges of access and disclosure, increasing the number of patients screened and extending care to those in need while reducing costs and the administrative burden associated with paper questionnaires. These technologies have already enabled researchers to conduct large-scale studies of cognitive processes, happiness and wellbeing [24, 37, 57, 74], yet have rarely been deployed in clinical practice. Recent HCI research has drawn attention to the challenges of long-term engagement [56, 71, 77], and some have expressed doubt with respect to the possibility of gathering self-reported data in daily life for periods longer than "2-4 weeks," after which "the quality of ESM responses is known to deteriorate" [87]. A recent review of experience sampling studies within Computer Science reported a median study length of 14 days and a median population group of 19 participants [11]. Although pregnancy entails a highly diverse patient group and demands a longitudinal perspective, little research has examined the feasibility of self-report technologies in the perinatal context.

3 METHODOLOGY

The Study Aims

The primary aim of this study was to assess the feasibility of a mobile application running on women's own devices for the repeated and longitudinal self-report of antenatal mood and depression. Secondary aims included comparing two distinct time-based sampling protocol designs, collecting momentary and retrospective reports of wellbeing in daily life, and examining the role of mobile devices as a means to address barriers to antenatal mental health screening. This entailed integrating mobile devices within NHS mental health screening pathways.

The Technology

This study involved the deployment of a technology, comprising mobile applications (Android & iOS) and a server for data storage, management and alert handling, within a public health service. This system, BrightSelf, was developed in collaboration with pregnant women, mothers and a variety of health professionals including midwives [7, 21]. This mobile application provides a platform for the self-report of psychological wellbeing in pregnancy, including retrospective reports in the form of the Edinburgh Postnatal Depression Scale (EPDS), and momentary reports in the form of visual analogue scales for mood, sleep, worry, enjoyment and energy as well as two questions concerning location and activity context (See Appendix A for video of the mobile app in use). The EPDS is a validated screening scale for depression which assesses feelings of guilt, sleep disturbance, anhedonia, self-harm and suicidal ideation present during the past 7 days [2, 20]. Scores between 10 and 12 points suggest a possible risk of depression, and 13 points or more, a probable risk. Additional features of the application include interactive visualisations of users' data, information regarding perinatal wellbeing and the study itself, contacts

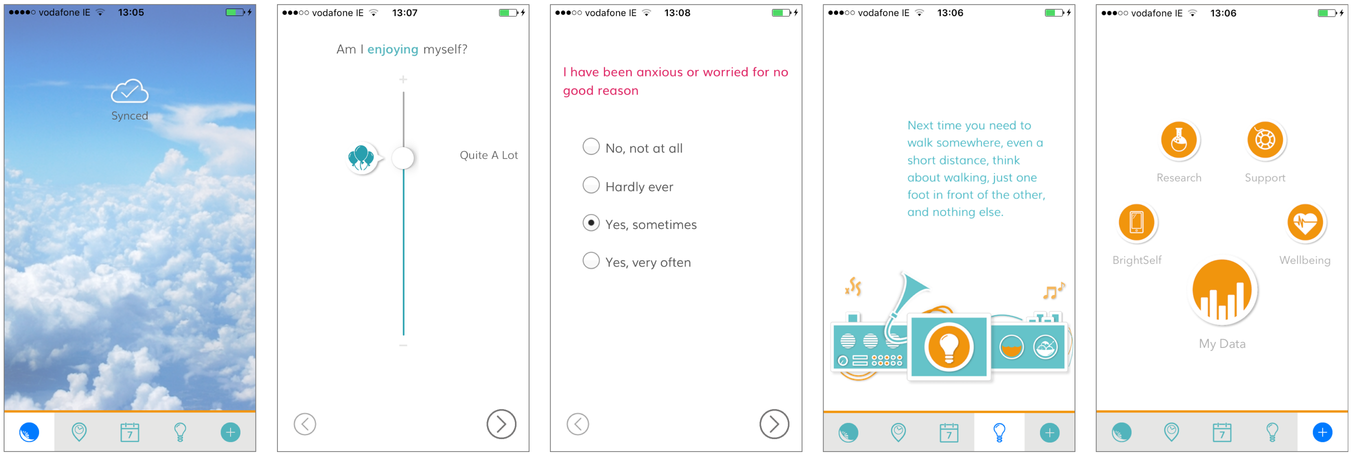

Figure 1: The BrightSelf Mobile Application | Home, EMA, EPDS, Ideas Machine and Extra Features Screens

for mental health support, and an 'ideas machine' designed to support longitudinal user engagement (See Fig 1).

The Study Design

Pregnant women attending NHS antenatal clinics across England were recruited to this study while attending their first 'booking' appointment between the 12th and 14th weeks of pregnancy [50]. Following informed consent, women completed a personal demographic survey, the Whooley Questions (a 2-item instrument which screens for depressed mood and anhedonia present during the past month) and the EPDS in-clinic. A member of the clinical care team then guided the participant through the process of downloading and installing the application onto their own device using an illustrated guide provided by the research team. When possible, installation took place within the clinic. However, internet services were not always available, in which case participants were provided a userID with which to install and register the app in their own time. Participants were randomly allocated to one of two arms of a study with a randomised controlled trial (RCT) design, permitting comparison of women's engagement between a retrospective assessment strategy and a retrospective plus momentary assessment strategy.

Participants in both arms were prompted to provide reports according to a 6-month burst protocol design and received their first notification 2 days following installation of the app (See Fig 2). In the *first condition*, participants were randomly prompted to provide a single EPDS report once per month between the hours of 17:00 and 21:00, 21 to 35 days apart, for a total of 6 notifications. In the *second condition*, participants had the option to complete both retrospective and momentary assessments. Monthly sampling periods consisted of 6 contiguous days of semi-random assessments, comprising a single EPDS prompt between 17:00 and 21:00 on day 1, followed by four days of momentary assessments 3 times per day (between 09:00-12:00, 13:00-16:00 and 17:00-20:00), and concluding with a final EPDS assessment between the hours of 17:00 and 21:00 on day 6. Prompting was repeated 21 to 35 days following the previous period, for a total of 6 assessment periods. Participants were free to provide reports, ignore or disable notifications, or delete the app at any time. No reminders or follow-up notifications were sent to participants, and no monetary incentives were provided. Participants' use of the application was logged throughout the study. Two weeks following their final notification, women received a survey designed to rule out potential confounds, triangulate a subjective measure of engagement and gauge the experience of use [50].

Duty of Care

This study design entailed significant ethical and medical responsibilities given the collection of data reflecting women's self-reported risk of depression, self-harm and suicide [7, 50]. Participants' encrypted and pseudo-anonymised data was synchronised with the research server using a secure connection throughout the study and a colour-coded alert was instantly communicated to the research team when any participant's data met EPDS risk criteria. Within 24-hours, and often almost immediately, the study coordinator contacted the designated member of the participant's clinical care team by phone and e-mail, who followed up directly with the participant. This study protocol was reviewed and approved by the National Research Ethics Service Committee South East Coast-Surrey on the 15th of April 2016 as a notice of substantial amendment to an original submission (9th July 2015) under the Research Ethics Committee (REC) reference 15/LO/0977. This study was sponsored by Imperial College London under the reference number 15IC2687 and included

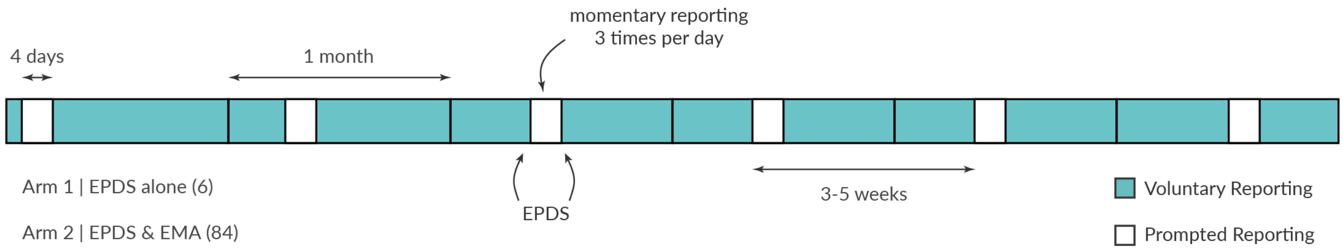

Figure 2: The Sampling Protocol Design

in the UK Clinical Research Network Study Portfolio under the Central Portfolio Management System number 19280.

4 RESULTS

This paper examines women's engagement with a mobile application for the self-report of wellbeing and depression in pregnancy.

The Study Population

Between April and September 2017, midwives at 14 NHS midwifery clinics across England recruited women to a study involving BrightSelf. Of the 355 women who consented to participate, 254 subsequently installed the app. 128 were randomly allocated to arm 1 (EPDS reporting only) and 126 to arm 2 (EPDS & EMA reporting) prior to app installation.

App Installation by Population & Wellbeing. Much of the promise of mobile technologies hinges upon their potential to extend care to at-risk and under-served groups. 45% of women participating in this study were experiencing their first pregnancy. 23% reported 'single' marital status, 45% a level of education below a university or college degree, 17% unemployment, and 21% an ethnicity other than White British. Examining the proportion of the study population[1] who installed the app[2] allows us to better understand the relationship between women's demographic characteristics and their tendency to install a mobile app for the self-report of psychological wellbeing in pregnancy.

72% of participants installed BrightSelf, including women of all ages (See Fig 3). 68% of these 254 women used iOS devices. There was no evidence of a statistically significant relationship between women's age (χ^2=9.8109, df=4, p=0.04374, n=348), marital status (χ^2=2.6289, df=3, p=0.4524, n=348), employment status (χ^2=10.385, df=4, p=0.03441, n=348), level of education (χ^2=9.867, df=4, p=0.04274, n=348), number of children (χ^2=3.6555, df=1, p=0.05589, n=348), tablet ownership (χ^2=0.0878, df=1, p=0.767, n=348) or a past diagnosis of depression (χ^2=3.114, df=1, p=0.07762, n=348) and installation of this application.[3] 32 women reported a past diagnosis of depression, 18 of whom installed BrightSelf. Of the 11 women who reported a previous pregnancy which did not reach full-term, only 2 did not install the app. Women of White British ethnicity were more likely to install BrightSelf than others (χ^2=10.58, df=1, p=0.001143, n=348), possibly due to site-specific or cultural factors.

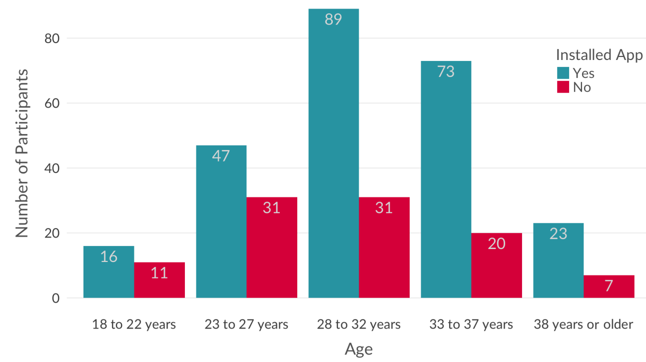

Figure 3: Participants by Age

Participants completed both the Whooley questions and EPDS at baseline. 27 women responded affirmatively to Whooley question one which queries depressed mood (of whom 14 installed BrightSelf) and 24 to question two which examines anhedonia (of whom 16 installed BrightSelf). There is no evidence of a statistically significant relationship between

[1] n=348. Of the 355 baseline data logs collected, 2 lacked participant ids and 5 featured empty data fields, likely due to a combination of poor internet connection in-clinic, bugs or crashes in the proprietary survey software, and errors made by recruiters.

[2] n=248. Of the 254 unique users who installed the app, 6 (3 arm 1, 3 arm 2) could not be linked to their demographic data provided at baseline, as described above. For consistency, we exclude the data of these 6 women from this preliminary analysis.

[3] We employ chi-square tests for independence given nominal dependent and independent variables. To protect from Type I errors, we conduct Bonferroni correction for multiple tests (8 tests on the single dependent variable of 'app installation') which yields a corrected alpha-value of 0.00625 for statistical significance in this case. While not all tests yielded statistically significant results according to this threshold, this analysis also serves to highlight population characteristics for additional future analyses.

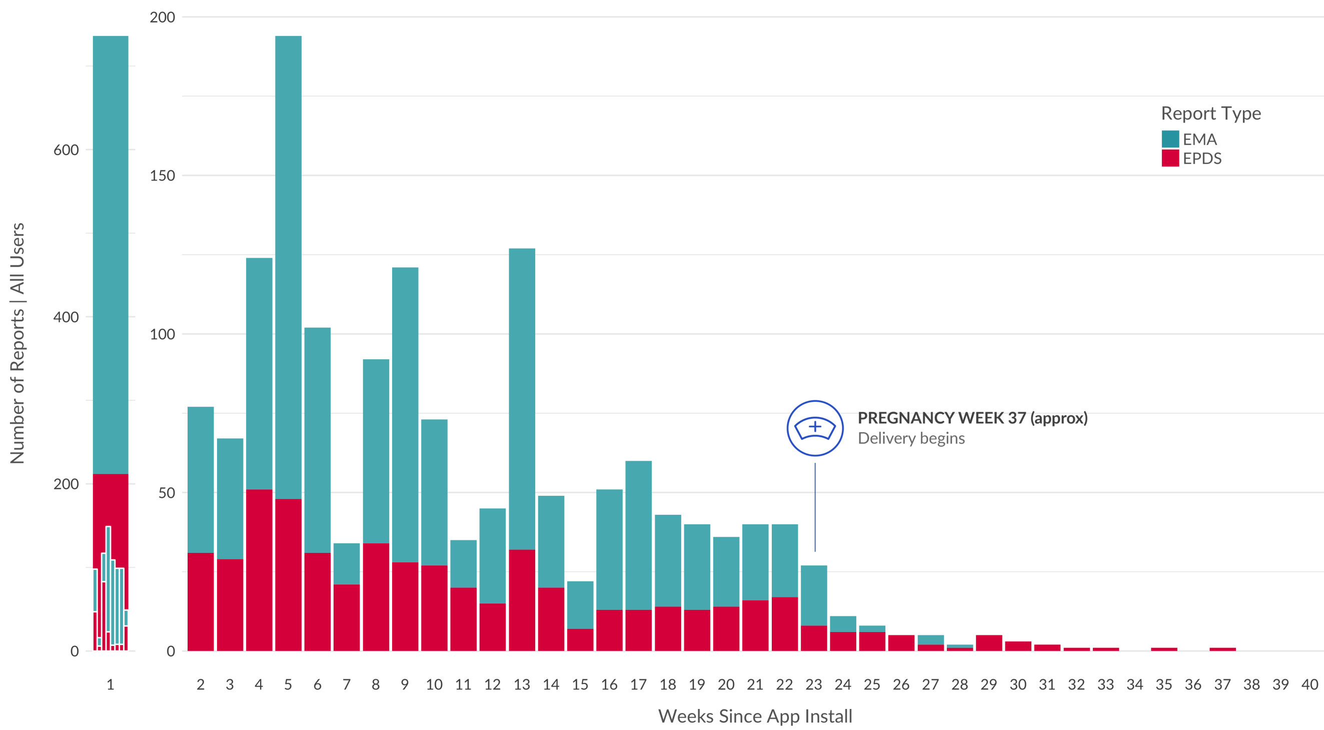

Figure 4: Reporting by Week | All Users

installation of BrightSelf and women's responses to Whooley questions one *(χ^2=0.38256, df=1, p=0.5362, n=348)* or two *(χ^2=2.2133, df=1, p=0.1368, n=348)*. The relationship between women's risk of depression according to the EPDS and installation of BrightSelf was also not statistically significant *(χ^2=1.7326, df=2, p=0.4205, n=348)*. 22 women scored between 10 and 12 points on the EPDS in-clinic, reflecting a possible risk of depression, and fifteen women scored 13 points or above, indicating a probable risk. Six women responded positively to EPDS question ten concerning self-harm ideation.

Evidence for Feasibility | Women's Engagement

The rich data captured by BrightSelf allows us to triangulate women's engagement from multiple perspectives.

Engagement as Self-Report. Over a 9 month period, women across England synchronised 2,280 reports using BrightSelf; 1,532 reports of their mood, energy, rest, enjoyment and worry in the moment and 748 retrospective EPDS reports.

The reporting trend reflects the characteristic context of pregnancy (See Fig 4).[4] The first week of use aligns with the 12th to 14th week of pregnancy, and the x-axis spans a period of 9 months. By the 23rd week of app use, many women have reached full-term. This pattern of interaction demonstrates the value of notifications in bringing women back to an app. The peaks at weeks 5, 9 and 13 are the result of women's overlapping reporting schedules. The potential for overlap decreases over time as the effects of randomisation increase. It is also worth noting however, just how many reports women provided during periods without any notifications, as evidenced by weeks 2 and 3 of app use in particular. Women in arm 1 synchronised 3.6 reports on average, and women in arm 2, 14.4, including those who provided no reports. Several women provided more than twice as many reports as they received notifications. Others logged their first report 20 or more weeks after installing the app only to generate an alert for a probable risk of depression.

Engagement as Protocol Adherence. Over the course of 6 months, women in arm 1 received 6 notifications. Women in arm 2 received 84. These protocols were chosen to invite

[4]n=254. Here, the first 8 days of use are represented separately as 'Week 1.' This is for two reasons; 1) The first day of use effectively represents a significant outlier given that all participants necessarily interact with the app on the day it is installed, and 2) During the first 8 days of use, the notification schedules of all participants are aligned. Subsequent notification periods are subject to randomisation by week within and between-subjects (a possible variance of 3 to 5 weeks at the second notification period increases to 10 to 15 by the sixth). This also has the effect of exaggerating the decline in use over time past the first week. Every subsequent week as graphed represents a period of 7 consecutive days since the morning of the day of installation.

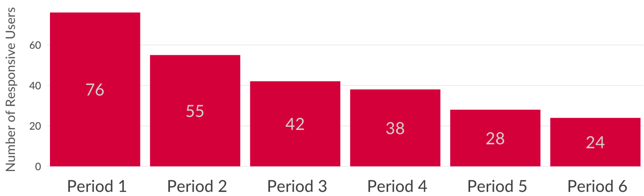

Figure 5: Protocol Adherence | Arm 1

	Period					
Report Type	1	2	3	4	5	6
EPDS Arm 1	76	55	42	38	28	24
EPDS Arm 2	88	41	31	25	13	17
EMA Arm 2	430	228	185	151	110	86
Combined Arm 2	518	269	216	176	123	103

Table 1: Reporting by Arm

and explore different reporting practices. Women could adhere to this schedule, provide reports whenever convenient, disable notifications or indeed delete the app. During a survey completed post-study, one woman admitted to disabling notifications and providing reports "when I felt like as often I can't respond when prompted as I'm at work." This participant's continued use of an app suggests the importance of supporting users' autonomy, even in the case of data collection to support research.

Women's adherence to these protocols also reflects their engagement over time. Figures 5 & 6 were produced by aligning the notification schedules of all participants and treating as a valid response any report which took place during the period between two notifications of the same type.[5] The spike in EMA reporting at the end of each reporting period is an artefact of this approach. Each EMA reporting period shown in Fig 6 is on average 4 hours in duration. Over a period of more than 6 months no period elapsed without a response from at least one participant. To facilitate comparisons between arms, we combine the total number of responses within each of the six reporting periods, as shown in Table 1.

In arm 1, 32% of those who provided a report during the first week of use responded to a final EPDS notification received 15 to 25 weeks later. 19% did so in the case of arm 2. However, women in arm 2 also provided 4 times as many reports as women in arm 1 on average. The addition of momentary reporting led to many more reports on average without greatly reducing the number of EPDS reports provided. These findings appear to mirror women's comments, made during the design phase, that they would be likely to complete EMA reports more often than the EPDS [21].

Engagement as Time Spent. The 'time spent' on app is another possible indicator of user engagement. On aggregate, women spent more than 52 hours interacting with BrightSelf, a total of 2,686 sessions with an average length of 1 minute and 24 seconds. Ancillary usage data allows us to examine exactly how women spent their time. Use of each of the main sections of the app was largely consistent over time, including time spent viewing data, accessing information and interacting with the Ideas Machine.

It should be noted that this application was not designed to maximise 'time on app.' Women's almost ubiquitous use of mobile apps in pregnancy results in an economy of attention in which designers can feel compelled to compete. As one woman states in the post-study survey, users may have spent more time on app had we incorporated more additional features, "I often have a flick through my app 'Baby Centre' looking at how the baby is developing, reading articles about pregnancy etc." We chose however to design for brief, simple and sporadic use based on input from women and professionals [21]. The average session duration reflects this efficiency of interaction and changed little per week, suggesting a consistent user experience. In the same post-study survey, 47% of women spontaneously described quick and easy interaction as the feature of the application which they most liked.

Women synchronised 812 uses of the Ideas Machine in total, and there is evidence of a statistically significant relationship between the number of reports provided and the use of this feature *(Spearman's correlation, S=2140100, $p<2.2e-16$, $\rho=0.6953$, n1=n2=254)*.[6] While correlation does not imply causation, women who used the Ideas Machine more frequently also provided more reports.

Engagement by Population & Wellbeing. Extending public health screening to those in need requires engaging patients regardless of their social status, cultural background or personal characteristics. In the case of the users of BrightSelf, no significant differences were found between women's age *(Kruskal-Wallis $\chi^2=4.0321$, df=4, p=0.4017, n=248)*, ethnicity *(Kruskal-Wallis $\chi^2=3.0946$, df=1, p=0.0786, n=248)*, marital status *(Kruskal-Wallis $\chi^2=2.5942$, df=3, p=0.4585, n=248)*, employment status *(Kruskal-Wallis $\chi^2=3.8395$, df=4, p=0.4282,*

[5] Arm1, n=120, Arm2, n=117. Unexpected behaviour in iOS10 led the data of 17 women (8 arm 1, 9 arm 2) to be returned in 12 rather than 24-hour time format, rendering their timestamped data unreliable to the hour. For the purposes of consistency we omit these users' data from this preliminary analysis.

[6] Spearman's rank-order correlation is the non-parametric version of the Pearson product-moment correlation, employed in the case of non-normal continuous dependent and independent variables.

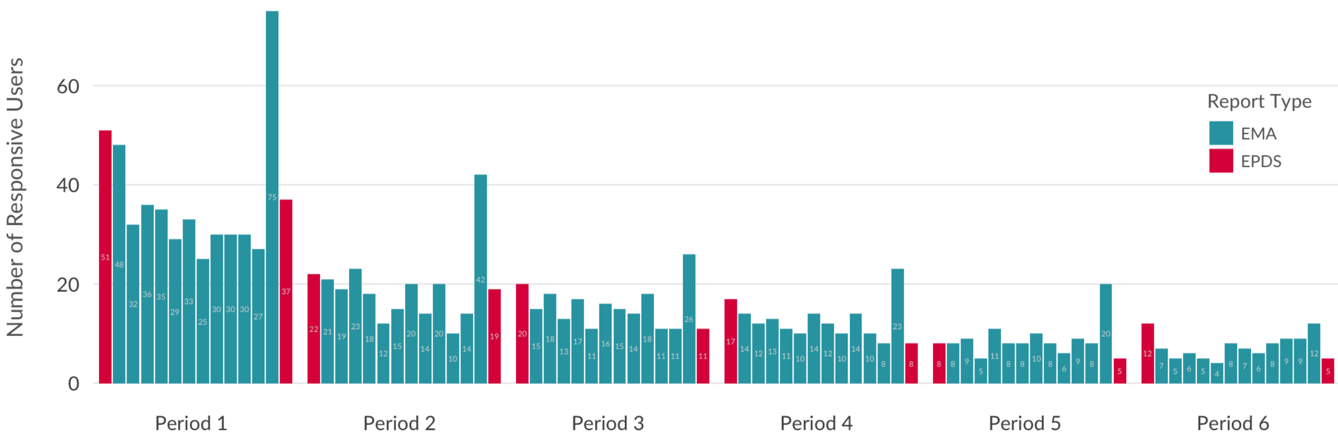

Figure 6: Protocol Adherence | Arm 2

n=248), level of education *(Kruskal-Wallis χ^2=3.9456, df=4, p=0.4134, n=248)*, tablet ownership *(Kruskal-Wallis χ^2=1.7774, df=1, p=0.1825, n=248)*, number of children *(Kruskal-Wallis χ^2=0.21977, df=1, p=0.6392, n=248)* or past diagnosis of depression *(Kruskal-Wallis χ^2=0.2469, df=1, p=0.6193, n=248)* and rates of reporting.[7] Women engaged with an application for the self-report of psychological wellbeing in pregnancy regardless of their demographic characteristics.

Similarly, no statistically significant difference was found between women meeting EPDS thresholds for no, possible and probable risks of depression at 12 to 14 weeks and their engagement in self-report during pregnancy in the case of arm 1 *(Kruskal-Wallis χ^2=2.3908, df=2, p=0.3026, n=125)*, arm 2 *(Kruskal-Wallis χ^2=1.9505, df=2, p=0.3771, n=123)* or all users combined *(Kruskal-Wallis χ^2=0.96303, df=2, p=0.6178, n=248)*. Nor was there a significant difference between women's baseline EPDS assessments and the total time spent using BrightSelf *(Kruskal-Wallis χ^2=1.16, df=2, p=0.5599, n=248)*. Women engaged with a self-report application in pregnancy regardless of their level of ill or wellbeing.

Distributions of women's use of the Ideas Machine according to their baseline EPDS assessments did differ significantly however *(Kruskal-Wallis χ^2=9.0126, df=2, p=0.01104, n=248)*. Women in greater distress were more likely to use this feature, suggesting it was perceived as useful by those in need.

Engagement & Wellbeing | Users in Distress. Of the 748 EPDS reports provided using BrightSelf, 71 met the threshold for possible (41) or probable (27) depression. Two EPDS reports at baseline reflected a risk of self-harm ideation. Fifteen did so through BrightSelf. Of the 39 unique women who registered alerts during the study, 7 provided EPDS scores of 10 at baseline, and 6 registered scores of 13 or above. 26 women therefore received support which they might not have acquired given their baseline EPDS reports alone.[8] 5% of the women who installed BrightSelf registered EPDS scores of 10 or above on paper at baseline. 16% did so through the use of the app in pregnancy.

Interestingly, of the 39 women who provided EPDS scores which resulted in contact from a midwife, 20 continued to provide EPDS reports and 13 provided at least one subsequent alert (See Fig 7). In the post-study survey, women who were contacted by their midwives following an alert spoke positively of the experience;

> "I had a call from a midwife to see if I was ok and if I needed any help. I think it should be made available to all pregnant women and not just for research purposes" Participant No. 294 | Arm 1

Subjective Engagement | Women's Reflections

> "I thought this was a brilliant app and an excellent research project. I don't know what factors have contributed to this but I feel a lot better after having my second child than I did my first" Participant No. 265 | Arm 2

62 women (25 arm 1, 37 arm 2) responded to a post-study survey seeking their experience of an app for the self-report of psychological wellbeing in pregnancy. The reported experiences of these women provide context for the preceding findings and allows us to rule out contingent effects related to the usability of the app. 92% of women agreed or strongly

[7]Reporting rates, total time spent on app and use of the Ideas Machine data do not follow normal distributions (as confirmed using quantile-quantile plots and the Shapiro-Wilk test). We therefore employ a Kruskal-Wallis test for independence as appropriate in the case of continuous dependent and ordinal independent variables with more than 2 groups.

[8]Two-thirds of these women presented no risk of depression according to either the EPDS or Whooley questions applied independently in-clinic. 54% presented no risk at baseline according to both methods combined.

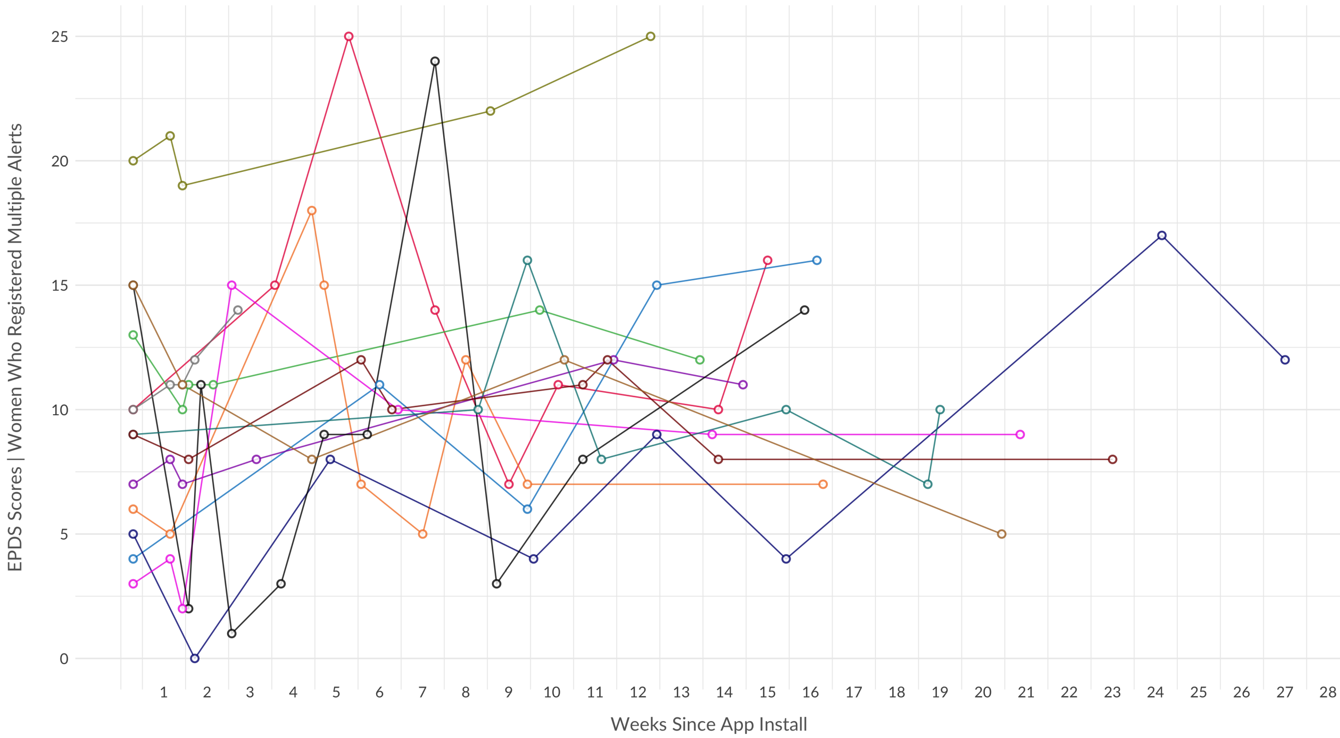

Figure 7: EPDS Scores | Participants Who Registered Multiple Alerts

agreed that the mobile app was easy to use. Similarly, 95% of women agreed or strongly agreed that they learned quickly to use the app. When asked whether the mobile app met their needs, 56% of participants responded affirmatively, with 26% undecided. 65% of respondents specified that they found the assessments useful, with 11% undecided.

We also surveyed women's responses to a number of open-ended questions. Responding to a question regarding their motivations for the use of a self-report app in pregnancy, 22 women described a desire to support research and 14 to help others; "[I] believe it could be helpful for people and wanted to be part of it." Women also described personal motivations; monitoring mental health or mood (n=7), supporting self-reflection (n=3), knowing someone who had struggled during pregnancy (n=3) and personal experience of a difficult pregnancy (n=1); "Sometimes I felt low and this help helped me to recognise those feelings." Several women commented that their motivations changed over time; "Initially it was because I am interested in research and mental health but once I started using the app I felt that it was useful to monitor how tired I felt which reminded me to take time for myself after a busy day at work."

This post-study survey also provided an opportunity to triangulate a qualitative assessment of women's engagement. 53% of women (with 20% undecided) rated the experience as engaging. 73% (with 10% undecided) reported that they would recommend BrightSelf to a friend, and 69% of respondents (with 15% undecided) stated that they would repeat the experience. When asked what they liked most about the experience of the mobile app, 29 women responded that it was quick and easy to use. 14 women stated that the app helped them to engage in mindful reflection; "It made me reflect on the week I had and think about how rational or irrational my emotions had been." Other responses included the sense that one was making a contribution (n=1), the Ideas Machine (n=2), privacy (n=1), unintrusive interaction (n=2), viewing data (n=4), and monitoring and support (n=4).

We also sought women's preferences with respect to the design of time-based sampling protocols (See Fig 8). 32% of respondents in arm 2, who received 84 notifications over a 6 month period, reported receiving notifications too frequently or much too frequently. 54% of respondents in arm 1, who received a single prompt once a month, reported that this was too infrequent or much too infrequent a schedule. When asked what they liked least about the experience of using the app, 25 women referred to the frequency of notifications, although women in both arms described notifications as both too frequent and infrequent. One woman expressed dismay that notifications stopped "towards the end of my pregnancy." Seven women cited the scales employed, requesting greater

Figure 8: Perceptions of Reporting Frequency | Post-Study

variety. One woman noted that at times of anxiety she felt less eager to use the app, and two requested the ability to reset their passwords. Women's comments concerning possible improvements to BrightSelf largely mirrored these responses. Finally, we asked participants in arm 2 whether they found it useful to compare the reports they provided in the moment with those made retrospectively (in the form of the EPDS). 52% (with 20% undecided) responded affirmatively.

There was no evidence of a statistically significant relationship between women's actual reporting practice and subjective ratings of engagement *(Kruskal-Wallis χ^2=6.4882, df=6, p=0.3708, n=61)*, willingness to repeat the experience *(Kruskal-Wallis χ^2=8.0513, df=6, p=0.2344, n=61)*, willingness to recommend the app to a friend *(Kruskal-Wallis χ^2=10.682, df=6, p=0.0988, n=61)*, ability of the app to meet their needs *(Kruskal-Wallis χ^2=12.258, df=6, p=0.0565, n=61)* or the perception of assessments as useful *(Kruskal-Wallis χ^2=2.9494, df=6, p=0.8152, n=61)*. This may be a consequence of sample size or self-selection bias among survey respondents or reflect the diverse patterns of reporting in which women engaged according to their needs. There was a statistically significant relationship between women's perception of the frequency of notifications and their reporting practice however *(Kruskal-Wallis χ^2=12.331, df=4, p=0.0151, n=61)*. Women who rated notifications as much too frequent and much too infrequent both provided fewer reports than other users.

5 DISCUSSION

This paper presents initial evidence for the feasibility of deploying mobile devices in antenatal mental health screening.

Engaging Patients in Self-Report

The feasibility of self-report technologies hinges in large part upon the engagement of users [10, 17, 18, 85]. This study design entailed a mixed-methods approach to its assessment. Over the course of 9 months, 254 women spent more than 52 hours interacting with a mobile application for the self-report of wellbeing and depression in pregnancy, providing 2,280 reports of their subjective experience without extrinsic incentives. These figures reflect the unique potential of mobile devices to engage patients in public health screening and data collection in daily life.

Extending Care to Under-Served and At-Risk Groups

Extending care to under-served and at-risk groups is a public health priority, and an essential step towards bridging the mental health treatment gap [67, 83]. Pregnant women are a particularly diverse population group, as demonstrated by the characteristics of those who took part in this study. However, women attending NHS antenatal clinics across England installed and engaged with a self-report application regardless of their age, marital status, employment status, level of education, number of children, tablet ownership, past diagnosis of depression, or risk of depression according to screening in-clinic. These findings represent initial evidence for the potential of mobile devices to extend care to a significant patient population.

Previous research has suggested that women of ethnic minority backgrounds in the UK may consider professional assistance less appropriate for the treatment of perinatal depression [75]. In this instance, ethnic minority women were found to be less likely to install a mobile application for the self-report of wellbeing and depression in pregnancy. However, once installed, women engaged with this application regardless of their ethnicity. This interesting finding suggests the need for future design efforts to focus on cultural differences with respect to the design of technologies for health and wellbeing, their introduction into clinical contexts, and the development of a broad portfolio of screening methods.

Gathering Longitudinal, Momentary and Retrospective Data

Assessing wellbeing during pregnancy necessitates both a longitudinal perspective and an understanding of subjective experience with respect to multiple time-frames [21, 22, 25, 29]. Little research has examined the self-report of mental health for periods longer than several weeks in any clinical context. During this study, however, many women engaged in reporting for periods of 6 months and longer, suggesting the potential of mobile devices to facilitate the remote screening and monitoring of mood and depression throughout the antenatal period. Women's motivations for doing so often reflected a desire to support both their own wellbeing and that of other women. Future analysis of this corpus of momentary and retrospective data will aim to advance our understanding of depression and wellbeing in pregnancy and inform the design of more effective public health screening programs.

This paper contributes evidence not just for the feasibility of these systems but for their appropriate design. 47% of survey respondents spontaneously described simple and efficient interaction as the characteristic of this self-report technology which they most liked. Several women commented that they found intrinsic value in the practice of self-report, mirroring motivations expressed by women during the design phase [21]. Women in greater distress used the Ideas Machine feature of the application more often. And, women who used this feature more often provided more reports.

The addition of momentary reporting in daily life led to many more reports on average without greatly reducing the number of EPDS reports provided by women, despite a heavier notification burden. A broader conception and understanding of ill and wellbeing has the potential to facilitate new opportunities for disclosure, care and research in the public health context. While the deployment of mobile devices for mental health screening may require the use of a validated scale such as the EPDS, these results suggest that the simultaneous collection of momentary data need not adversely affect completion of the primary clinical measure.

Finally, women who rated notifications as much too frequent or infrequent both provided fewer reports than other users. This finding highlights the importance of appropriate sampling protocol design, which this combined analysis of notification settings, reporting practice and the subjective experience of self-report provides initial evidence for.

Overcoming Stigma and Supporting Disclosure

> "Everyone should have access to the app as soon as they find out their [sic] pregnant, great way of communicating, especially for those less inclined to talk to anyone" Participant No. 349 | Arm 1

Mental health related stigma is one of the primary barriers to care during pregnancy [28, 35, 39, 67, 83]. During this study, 39 women disclosed a risk of depression, self-harm or suicide and received immediate midwife support. Two-thirds of participants who received support in this way registered no risk of depression according to standard screening methods employed in-clinic at baseline. These figures suggest the potential of a mobile application deployed on women's personal devices to overcome stigma and support disclosure, facilitating care and support for those in need. Women at risk were identified during pregnancy at a rate similar to that of postnatal depression diagnosis nationally, reflecting the potential advantage of this opportunity to provide early intervention and support [5, 48].

Fostering Trust Between Patients and Professionals

Introducing a new source of patient data into any clinical context entails significant ethical and medical responsibilities. Designers of health and wellbeing technologies must attend to their potential to overburden professionals, dehumanise care, medicalise pregnancy, fail to manage high-risk cases or adversely shape the criteria by which decisions are made [7, 33]. Future configurations of public health systems and services may take many forms. Clinical interfaces may be required, for example, to support data management, sense-making, risk-assessment and patient-provider relationships at a national scale [15, 38, 54].

The integration of personal information and public health implies a new form of clinical practice, characterised by patient participation and respect for the 'co-creation' of health and wellbeing. This is a highly complex context for research and design, combining the personal significance of wellbeing during pregnancy, the real-world and long-term use of technology by an at-risk user group, a public health system's need to efficiently distribute resources, midwives' work practices, social expectations, societal stigma, and researchers' motivations for data-collection. Designing any technology capable of respecting these constraints is a 'wicked problem' as first defined by Rittel & Webber in 1973 [69].

Responding to these design challenges requires attending to the needs, values, expectations and concerns of users. Public health researchers, pregnant women and a variety of health professionals including midwives were involved in the development of the system deployed during this study. The women who received support using this technology spoke positively of the experience and many continued to engage in reporting thereafter. These findings further suggest the potential of mobile devices to efficiently direct care towards those in need at little extra cost while fostering trust between women and midwives.

6 CONCLUSIONS

This paper contributes initial evidence for the appropriate design and feasibility of mobile applications to engage women in the self-report of wellbeing and depression during pregnancy, extend care to under-served and at-risk populations, enable longitudinal, momentary and retrospective data collection, overcome stigma, support disclosure, and foster trust between patients and health professionals.

ACKNOWLEDGMENTS

This research is supported by Science Foundation Ireland through Grant 12/CE/I2267 to the Adapt Centre and a National Institute for Health Research (NIHR) Imperial Biomedical Research Centre award through the Population Health Theme. We would also like to thank the many women and healthcare professionals across the UK who so willingly gave of their time, experience and expertise to make this work possible.

REFERENCES

[1] Alexander T Adams, Elizabeth L Murnane, Phil Adams, Michael Elfenbein, Pamara F Chang, Shruti Sannon, Geri Gay, and Tanzeem Choudhury. 2018. Keppi: A Tangible User Interface for Self-Reporting Pain. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*. ACM, 502.

[2] Lucy J Allbaugh, Sheila M Marcus, Eleanor C Ford, and Heather A Flynn. 2015. Development of a screening and recruitment registry to facilitate perinatal depression research in obstetrics settings in the USA. *International Journal of Gynecology & Obstetrics* 128, 3 (2015), 260–263.

[3] Nazanin Andalibi and Andrea Forte. 2018. Announcing Pregnancy Loss on Facebook: A Decision-Making Framework for Stigmatized Disclosures on Identified Social Network Sites. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*. ACM, 158.

[4] Michael F Armey, Janis H Crowther, and Ivan W Miller. 2011. Changes in ecological momentary assessment reported affect associated with episodes of nonsuicidal self-injury. *Behavior therapy* 42, 4 (2011), 579–588.

[5] Patients Association et al. 2011. Postnatal Depression Services: An Investigation into NHS Service Provision. (2011).

[6] Marc S Atkins and Stacy L Frazier. 2011. Expanding the Toolkit or Changing the Paradigm: Are We Ready for a Public Health Approach to Mental Health? *Perspectives on Psychological Science* 6, 5 (2011), 483–487.

[7] Marguerite Barry, Kevin Doherty, José Marcano-Belisario, Josip Car, Cecily Morrison, and Gavin Doherty. 2017. mHealth for Maternal Mental Health: Everyday Wisdom in Ethical Design. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*. ACM, 2708–2756.

[8] Amna Batool, Samia Razaq, Maham Javaid, Beenish Fatima, and Kentaro Toyama. 2017. Maternal Complications: Nuances in Mobile Interventions for Maternal Health in Urban Pakistan. In *Proceedings of the Ninth International Conference on Information and Communication Technologies and Development (ICTD '17)*. ACM, 3:1–3:12.

[9] Annette Bauer, Michael Parsonage, Martin Knapp, Valentina Iemmi, and Bayo Adelaja. 2014. The costs of perinatal mental health problems. (2014).

[10] Daniel J Beal and Howard M Weiss. 2003. Methods of ecological momentary assessment in organizational research. *Organizational Research Methods* 6, 4 (2003), 440–464.

[11] Niels Van Berkel, Denzil Ferreira, and Vassilis Kostakos. 2017. The Experience Sampling Method on Mobile Devices. *ACM Computing Surveys (CSUR)* 50, 6 (2017), 93.

[12] Boots Family Trust. 2013. Perinatal Mental Health: Experiences of Women and Health Professionals. *Boots Family Trust, London* (2013).

[13] Elizabeth C Braithwaite, Susannah E Murphy, and Paul G Ramchandani. 2016. Effects of prenatal depressive symptoms on maternal and infant cortisol reactivity. *Archives of women's mental health* 19, 4 (2016), 581–590.

[14] Joan E Broderick, Joseph E Schwartz, Gregory Vikingstad, Michelle Pribbernow, Steven Grossman, and Arthur A Stone. 2008. The accuracy of pain and fatigue items across different reporting periods. *Pain* 139, 1 (2008), 146–157.

[15] Chris M Buonocore, Rosemary A Rocchio, Alfonso Roman, Christine E King, and Majid Sarrafzadeh. 2017. Wireless sensor-dependent ecological momentary assessment for pediatric asthma mhealth applications. In *Proceedings of the Second IEEE/ACM International Conference on Connected Health: Applications, Systems and Engineering Technologies*. IEEE Press, 137–146.

[16] Nancy Byatt, Kathleen Biebel, Liz Friedman, Gifty Debordes-Jackson, Douglas Ziedonis, and Lori Pbert. 2013. Patient's views on depression care in obstetric settings: how do they compare to the views of perinatal health care professionals? *General hospital psychiatry* 35, 6 (2013), 598–604.

[17] Mauro Cherubini and Nuria Oliver. 2009. A refined experience sampling method to capture mobile user experience. *International Workshop of Mobile User Experience Research, CHI'2009* (2009).

[18] Tamlin Conner Christensen, Lisa Feldman Barrett, Eliza Bliss-Moreau, Kirsten Lebo, and Cynthia Kaschub. 2003. A practical guide to experience-sampling procedures. *Journal of Happiness Studies* 4, 1 (2003), 53–78.

[19] Tamlin S Conner and Lisa Feldman Barrett. 2012. Trends in ambulatory self-report: the role of momentary experience in psychosomatic medicine. *Psychosomatic medicine* 74, 4 (2012), 327.

[20] John L Cox, Jeni M Holden, and Ruth Sagovsky. 1987. Detection of postnatal depression. Development of the 10-item Edinburgh Postnatal Depression Scale. *The British journal of psychiatry* 150, 6 (1987), 782–786.

[21] Kevin Doherty, Marguerite Barry, José Marcano-Belisario, Bérenger Arnaud, Cecily Morrison, Josip Car, and Gavin Doherty. 2018. A Mobile App for the Self-Report of Psychological Well-Being During Pregnancy (BrightSelf): Qualitative Design Study. *JMIR mental health* 5, 4 (2018).

[22] Kevin Doherty and Gavin Doherty. 2018. The construal of experience in HCI: Understanding self-reports. *International Journal of Human-Computer Studies* 110 (2018), 63–74.

[23] Kevin Doherty and Gavin Doherty. 2018. Engagement in HCI: Conception, Theory and Measurement. *ACM Computing Surveys (CSUR)* 51, 5 (2018), 99:1–99:39.

[24] Stephane Dufau, Jon Andoni Duñabeitia, Carmen Moret-Tatay, Aileen McGonigal, David Peeters, F-Xavier Alario, David A Balota, Marc Brysbaert, Manuel Carreiras, Ludovic Ferrand, et al. 2011. Smart phone, smart science: how the use of smartphones can revolutionize research in cognitive science. *PloS one* 6, 9 (2011).

[25] Ulrich W Ebner-Priemer and Timothy J Trull. 2009. Ecological momentary assessment of mood disorders and mood dysregulation. *Psychological Assessment* 21, 4 (2009), 463.

[26] Daniel A Epstein, Nicole B Lee, Jennifer H Kang, Elena Agapie, Jessica Schroeder, Laura R Pina, James Fogarty, Julie A Kientz, and Sean Munson. 2017. Examining Menstrual Tracking to Inform the Design of Personal Informatics Tools. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*. ACM, 6876–6888.

[27] Yuan Gao, Xinying Li, Yu-Hsuan Lin, Xin Liu, and Lin Pang. 2014. Nuwa: Enhancing the Pregnancy Experience for Expectant Parents. In *CHI '14 Extended Abstracts on Human Factors in Computing Systems (CHI EA '14)*. ACM, New York, NY, USA, 257–262.

[28] Janice H Goodman. 2009. Women's attitudes, preferences, and perceived barriers to treatment for perinatal depression. *Birth* 36, 1 (2009), 60–69.

[29] Xinning Gui, Yu Chen, Yubo Kou, Katie Pine, and Yunan Chen. 2017. Investigating Support Seeking from Peers for Pregnancy in Online Health Communities. *Proc. ACM Hum.-Comput. Interact.* 1, CSCW (2017), 50:1–50:19.

[30] Gillian R Hayes, Karen G Cheng, Sen H Hirano, Karen P Tang, Marni S Nagel, and Dianne E Baker. 2014. Estrellita: A mobile capture and access tool for the support of preterm infants and their caregivers. *ACM Transactions on Computer-Human Interaction (TOCHI)* 21, 3 (2014), 19.

[31] Joel M Hektner, Jennifer A Schmidt, and Mihaly Csikszentmihalyi. 2007. *Experience sampling method: Measuring the quality of everyday life*. Sage.

[32] Jonathan Heron, Thomas G O'Connor, Jonathan Evans, Jean Golding, Vivette Glover, ALSPAC Study Team, et al. 2004. The course of anxiety

and depression through pregnancy and the postpartum in a community sample. *Journal of affective disorders* 80, 1 (2004), 65–73.
[33] Alexandra Hillman. 2014. 'Why must I wait?' The performance of legitimacy in a hospital emergency department. *Sociology of health & illness* 36, 4 (2014), 485–499.
[34] Stefan E Hormuth. 1986. The sampling of experiences in situ. *Journal of personality* 54, 1 (1986), 262–293.
[35] Louise M Howard, Emma Molyneaux, Cindy-Lee Dennis, Tamsen Rochat, Alan Stein, and Jeannette Milgrom. 2014. Non-psychotic mental disorders in the perinatal period. *The Lancet* 384, 9956 (2014), 1775–1788.
[36] Mary Hui, Christine Ly, and Carman Neustaedter. 2012. MammiBelli: Sharing Baby Activity Levels Between Expectant Mothers and Their Intimate Social Groups. In *CHI '12 Extended Abstracts on Human Factors in Computing Systems (CHI EA '12)*. ACM, New York, NY, USA, 1649–1654.
[37] Matthew A Killingsworth and Daniel T Gilbert. 2010. A wandering mind is an unhappy mind. *Science* 330, 6006 (2010), 932–932.
[38] Yoojung Kim, Eunyoung Heo, Hyunjeong Lee, Sookyoung Ji, Jueun Choi, Jeong-Whun Kim, Joongseek Lee, and Sooyoung Yoo. 2017. Prescribing 10,000 steps like aspirin: designing a novel interface for data-driven medical consultations. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*. ACM, 5787–5799.
[39] Dawn Kingston, Marie-Paule Austin, Maureen Heaman, Sheila McDonald, Gerri Lasiuk, Wendy Sword, Rebecca Giallo, Kathy Hegadoren, Lydia Vermeyden, Sander Veldhuyzen van Zanten, et al. 2015. Barriers and facilitators of mental health screening in pregnancy. *Journal of affective disorders* 186 (2015), 350–357.
[40] Michael T Kinsella and Catherine Monk. 2009. Impact of maternal stress, depression & anxiety on fetal neurobehavioral development. *Clinical obstetrics and gynecology* 52, 3 (2009), 425.
[41] Marian Knight, Derek Tuffnell, Sara Kenyon, Judy Shakespeare, Ron Gray, and Jennifer J Kurinczuk. 2016. Saving lives, improving mothers care: Surveillance of maternal deaths in the UK 2012-14 and lessons learned to inform maternity care from the UK and Ireland. Confidential enquiries into maternal deaths and morbidity 2009-14. *Oxford: National Perinatal Epidemiology Unit, University of Oxford* (2016).
[42] Andreas Koop. 2002. Lessons learned from 16 years usage of mobile computers in clinical trials. *Mobile Computing in Medicine* 15 (2002), 9–24.
[43] Jennifer L Kraschnewski, Cynthia H Chuang, Erika S Poole, Tamara Peyton, Ian Blubaugh, Jaimey Pauli, Alyssa Feher, and Madhu Reddy. 2014. Paging "Dr. Google": does technology fill the gap created by the prenatal care visit structure? Qualitative focus group study with pregnant women. *Journal of Medical Internet Research* 16, 6 (2014).
[44] Giordano Lanzola, MG Ginardi, A Mazzanti, and Silvana Quaglini. 2014. Gquest: Modeling patient questionnaires and administering them through a mobile platform application. *Computer methods and programs in biomedicine* 117, 2 (2014), 277–291.
[45] Jakob Eg Larsen, Thomas Blomseth Christiansen, and Kasper Eskelund. 2017. Fostering bilateral patient-clinician engagement in active self-tracking of subjective experience. In *Proceedings of the 11th EAI International Conference on Pervasive Computing Technologies for Healthcare*. ACM, 427–430.
[46] David Lazer, Alex Sandy Pentland, Lada Adamic, Sinan Aral, Albert Laszlo Barabasi, Devon Brewer, Nicholas Christakis, Noshir Contractor, James Fowler, Myron Gutmann, et al. 2009. Life in the network: the coming age of computational social science. *Science (New York, NY)* 323, 5915 (2009), 721.
[47] Tzu-I Lee, Yih-Harn Chiang, Jiayi Guo, Mu-Tsz Chen, and Yue Chen. 2016. Dot-it: Managing Nausea and Vomiting for A Peaceful Pregnancy with Personal Pattern Exploration. In *Proceedings of the 2016 CHI Conference Extended Abstracts on Human Factors in Computing Systems*. ACM, 20–25.
[48] Bronwyn Leigh and Jeannette Milgrom. 2008. Risk factors for antenatal depression, postnatal depression and parenting stress. *BMC psychiatry* 8, 1 (2008), 24.
[49] Juan F Maestre, Haley MacLeod, Ciabhan L Connelly, Julia C Dunbar, Jordan Beck, Katie A Siek, and Patrick C Shih. 2018. Defining through expansion: conducting asynchronous remote communities (arc) research with stigmatized groups. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*. ACM, 557.
[50] José Salvador Marcano-Belisario, Kevin Doherty, John O'Donoghue, Paul Ramchandani, Azeem Majeed, Gavin Doherty, Cecily Morrison, and Josip Car. 2017. A bespoke mobile application for the longitudinal assessment of depression and mood during pregnancy: protocol of a feasibility study. *BMJ Open* 7, 5 (2017).
[51] Jurgen Margraf, C Barr Taylor, Anke Ehlers, Walton T Roth, and W Stewart Agras. 1987. Panic Attacks in the Natural Environment. *The Journal of Nervous and Mental Disease* 175, 9 (1987), 558–565.
[52] Mark Matthews and Gavin Doherty. 2011. In the mood: engaging teenagers in psychotherapy using mobile phones. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. ACM, 2947–2956.
[53] Mark Matthews, Stephen Voida, Saeed Abdullah, Gavin Doherty, Tanzeem Choudhury, Sangha Im, and Geri Gay. 2015. In Situ Design for Mental Illness: Considering the Pathology of Bipolar Disorder in mHealth Design. In *Proceedings of the 17th International Conference on Human-Computer Interaction with Mobile Devices and Services*. ACM, 86–97.
[54] Helena M Mentis, Anita Komlodi, Katrina Schrader, Michael Phipps, Ann Gruber-Baldini, Karen Yarbrough, and Lisa Shulman. 2017. Crafting a view of self-tracking data in the clinical visit. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*. ACM, 5800–5812.
[55] Robin Mermelstein, Donald Hedeker, Brian Flay, and Saul Shiffman. 2007. Real-time data capture and adolescent cigarette smoking: moods and smoking. *The Science of Real-Time Data Capture: Self-Report in Health Research* (2007), 117–135.
[56] Jochen Meyer, Daniel Epstein, Parisa Eslambolchilar, Judy Kay, and Lie Ming Tang. 2018. A Short Workshop on Next Steps Towards Long Term Self Tracking. In *Extended Abstracts of the 2018 CHI Conference on Human Factors in Computing Systems*. ACM, W05:1–W05:8.
[57] Geoffrey Miller. 2012. The smartphone psychology manifesto. *Perspectives on Psychological Science* 7, 3 (2012), 221–237.
[58] Donna Moore, Susan Ayers, and Nicholas Drey. 2016. A Thematic Analysis of Stigma and Disclosure for Perinatal Depression on an Online Forum. *JMIR Mental Health* 3, 2 (19 May 2016).
[59] Sriraam Natarajan, Annu Prabhakar, Nandini Ramanan, Anna Baglione, Kay Connelly, and Katie Siek. 2017. Boosting for postpartum depression prediction. In *Proceedings of the Second IEEE/ACM International Conference on Connected Health: Applications, Systems and Engineering Technologies*. IEEE Press, 232–240.
[60] Mikael Palmblad and Brian Tiplady. 2004. Electronic diaries and questionnaires: designing user interfaces that are easy for all patients to use. *Quality of Life research* 13, 7 (2004), 1199–1207.
[61] Gaurav Paruthi, Shriti Raj, Ankita Gupta, Chuan-Che Huang, Yung-Ju Chang, and Mark W Newman. 2017. HEED: situated and distributed interactive devices for self-reporting. In *Proceedings of the 2017 ACM International Joint Conference on Pervasive and Ubiquitous Computing and Proceedings of the 2017 ACM International Symposium on Wearable Computers*. ACM, 181–184.
[62] Rebecca M Pearson, Jonathan Evans, Daphne Kounali, Glyn Lewis, Jon Heron, Paul G Ramchandani, Tom G O'Connor, and Alan Stein.

2013. Maternal depression during pregnancy and the postnatal period: risks and possible mechanisms for offspring depression at age 18 years. *JAMA psychiatry* 70, 12 (2013), 1312–1319.

[63] Trevor Perrier, Nicola Dell, Brian DeRenzi, Richard Anderson, John Kinuthia, Jennifer Unger, and Grace John-Stewart. 2015. Engaging Pregnant Women in Kenya with a Hybrid Computer-Human SMS Communication System. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems (CHI '15)*. ACM, New York, NY, USA, 1429–1438.

[64] Tamara Peyton, Erika Poole, Madhu Reddy, Jennifer Kraschnewski, and Cynthia Chuang. 2014. "Every Pregnancy is Different": Designing mHealth for the Pregnancy Ecology. In *Proceedings of the 2014 Conference on Designing Interactive Systems (DIS '14)*. ACM, New York, NY, USA, 577–586.

[65] Tamara Peyton, Erika Poole, Madhu Reddy, Jennifer Kraschnewski, and Cynthia Chuang. 2014. Information, Sharing and Support in Pregnancy: Addressing Needs for mHealth Design. In *Proceedings of the Companion Publication of the 17th ACM Conference on Computer Supported Cooperative Work & Social Computing (CSCW Companion '14)*. ACM, New York, NY, USA, 213–216.

[66] Annu Sible Prabhakar, Lucia Guerra-Reyes, Anne Effron, Vanessa M. Kleinschmidt, Maggie Driscoll, Charles Peters, Vanessa Pereira, Majdah Alshehri, Tom Ongwere, and Katie A. Siek. 2017. "Let Me Know if You Need Anything": Support Realities of New Mothers. In *Proceedings of the 11th EAI International Conference on Pervasive Computing Technologies for Healthcare (PervasiveHealth '17)*. ACM, New York, NY, USA, 31–40.

[67] NHS Improving Quality. 2015. Improving Access to Perinatal Mental Health Services in England - A Review. (2015).

[68] Mashfiqui Rabbi, Meredith Philyaw-Kotov, Jinseok Lee, Anthony Mansour, Laura Dent, Xiaolei Wang, Rebecca Cunningham, Erin Bonar, Inbal Nahum-Shani, Predrag Klasnja, et al. 2017. SARA: a mobile app to engage users in health data collection. In *Proceedings of the 2017 ACM International Joint Conference on Pervasive and Ubiquitous Computing and Proceedings of the 2017 ACM International Symposium on Wearable Computers*. ACM, 781–789.

[69] Horst WJ Rittel and Melvin M Webber. 1973. Dilemmas in a general theory of planning. *Policy sciences* 4, 2 (1973), 155–169.

[70] Umaira Uzma Sajjad and Suleman Shahid. 2016. Baby+: A Mobile Application to Support Pregnant Women in Pakistan. In *Proceedings of the 18th International Conference on Human-Computer Interaction with Mobile Devices and Services Adjunct (MobileHCI '16)*. ACM, New York, NY, USA, 667–674.

[71] Kavous Salehzadeh Niksirat, Sayan Sarcar, Huatong Sun, Effie LC Law, Torkil Clemmensen, Jeffrey Bardzell, Antti Oulasvirta, Chaklam Silpasuwanchai, Ann Light, and Xiangshi Ren. 2018. Approaching Engagement towards Human-Engaged Computing. In *Extended Abstracts of the 2018 CHI Conference on Human Factors in Computing Systems*. ACM, SIG14.

[72] Philip S Santangelo, Ulrich W Ebner-Priemer, and Timothy J Trull. 2013. Experience Sampling Methods in Clinical Psychology. *The Oxford Handbook of Research Strategies for Clinical Psychology* (2013), 188.

[73] Christie Napa Scollon, Chu-Kim Prieto, and Ed Diener. 2009. Experience sampling: promises and pitfalls, strength and weaknesses. In *Assessing well-being*. Springer, 157–180.

[74] Sandra Servia-Rodríguez, Kiran K Rachuri, Cecilia Mascolo, Peter J Rentfrow, Neal Lathia, and Gillian M Sandstrom. 2017. Mobile sensing at the service of mental well-being: a large-scale longitudinal study. In *Proceedings of the 26th International Conference on World Wide Web*. International World Wide Web Conferences Steering Committee, 103–112.

[75] Arshi Shaikh and Carol Kauppi. 2015. Postpartum depression: Deconstructing the label through a social constructionist lens. *Social Work in Mental Health* 13, 5 (2015), 459–480.

[76] Saul Shiffman, Arthur A Stone, and Michael R Hufford. 2008. Ecological momentary assessment. *Annu. Rev. Clin. Psychol.* 4 (2008), 1–32.

[77] Wally Smith, Bernd Ploderer, Greg Wadley, Sarah Webber, and Ron Borland. 2017. Trajectories of engagement and disengagement with a story-based smoking cessation app. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*. ACM, 3045–3056.

[78] Wally Smith, Greg Wadley, Oliver Daly, Marianne Webb, Jo Hughson, John Hajek, Anna Parker, Robyn Woodward-Kron, and David Story. 2017. Designing an app for pregnancy care for a culturally and linguistically diverse community. In *Proceedings of the 29th Australian Conference on Computer-Human Interaction*. ACM, 337–346.

[79] Marika B Solhan, Timothy J Trull, Seungmin Jahng, and Phillip K Wood. 2009. Clinical assessment of affective instability: comparing EMA indices, questionnaire reports, and retrospective recall. *Psychological Assessment* 21, 3 (2009), 425.

[80] Donna E Stewart. 2011. Depression during pregnancy. *New England Journal of Medicine* 365, 17 (2011), 1605–1611.

[81] Arthur A Stone, Joseph E Schwartz, John M Neale, Saul Shiffman, Christine A Marco, Mary Hickcox, Jean Paty, Laura S Porter, and Laura J Cruise. 1998. A comparison of coping assessed by ecological momentary assessment and retrospective recall. *Journal of personality and social psychology* 74, 6 (1998), 1670.

[82] Hyewon Suh, John R Porter, Alexis Hiniker, and Julie A Kientz. 2014. @BabySteps: design and evaluation of a system for using twitter for tracking children's developmental milestones. In *Proceedings of the 32nd annual ACM conference on Human factors in computing systems*. ACM, 2279–2288.

[83] Sara Thurgood, Daniel M Avery, and Lloyda Williamson. 2009. Postpartum depression (PPD). *American Journal of Clinical Medicine* 6, 2 (2009), 17–22.

[84] Phil Topham, Praminda Caleb-Solly, Paul Matthews, Andy Farmer, and Chris Mash. 2015. Mental Health App Design: A Journey From Concept to Completion. In *Proceedings of the 17th International Conference on Human-Computer Interaction with Mobile Devices and Services Adjunct*. ACM, 582–591.

[85] John Torous, Steven Richard Chan, Shih Yee-Marie Tan, Jacob Behrens, Ian Mathew, Erich J Conrad, Ladson Hinton, Peter Yellowlees, and Matcheri Keshavan. 2014. Patient smartphone ownership and interest in mobile apps to monitor symptoms of mental health conditions: a survey in four geographically distinct psychiatric clinics. *JMIR Mental Health* 1, 1 (2014).

[86] Nadia Tripp, Kirsten Hainey, Anthony Liu, Alison Poulton, Michael Peek, Jinman Kim, and Ralph Nanan. 2014. An emerging model of maternity care: smartphone, midwife, doctor? *Women and Birth* 27, 1 (2014), 64–67.

[87] Niels Van Berkel, Jorge Goncalves, Simo Hosio, and Vassilis Kostakos. 2017. Gamification of mobile experience sampling improves data quality and quantity. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 1, 3 (2017), 107.

[88] Julio Vega, Samuel Couth, Ellen Poliakoff, Sonja Kotz, Matthew Sullivan, Caroline Jay, Markel Vigo, and Simon Harper. 2018. Back to Analogue: Self-Reporting for Parkinson's Disease. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*. ACM, 74.

[89] Chelsea-Joy Wardle, Mitchell Green, Christine Wanjiru Mburu, and Melissa Densmore. 2018. Exploring Co-design with Breastfeeding Mothers. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*. ACM, 482.

[90] Max S. Wenger, Jarad Bell, Peter McEvoy, Cherie Yamaguchi, and Auriana Shokrpour. 2014. Bloom: Fostering Healthy and Peaceful Pregnancies with Personal Analytics. In *CHI '14 Extended Abstracts on Human Factors in Computing Systems (CHI EA '14)*. ACM, New York, NY, USA, 245–250.

[91] M Wichers, CJP Simons, IMA Kramer, JA Hartmann, C Lothmann, I Myin-Germeys, AL van Bemmel, F Peeters, Ph Delespaul, and J van Os. 2011. Momentary assessment technology as a tool to help patients with depression help themselves. *Acta Psychiatrica Scandinavica* 124, 4 (2011), 262–272.

[92] Anouk Wierckx, Suleman Shahid, and Abdullah Al Mahmud. 2014. Babywijzer: An Application to Support Women During Their Pregnancy. In *Proceedings of the Extended Abstracts of the 32Nd Annual ACM Conference on Human Factors in Computing Systems (CHI EA '14)*. ACM, New York, NY, USA, 1333–1338.

[93] Tracy K Witte, Kathleen K Fitzpatrick, Thomas E Joiner, and Norman Bradley Schmidt. 2005. Variability in suicidal ideation: A better predictor of suicide attempts than intensity or duration of ideation? *Journal of Affective Disorders* 88, 2 (2005), 131–136.