**TITLE PAGE The Bumps and BaBies Longitudinal Study (BaBBLeS): a multi-site cohort study of first-time mothers to evaluate the effectiveness of the Baby Buddy app**

*¥Toity Deave; e-mail: Toity.deave@uwe.ac.uk,

*Samuel Ginja; e-mail: s.ginja@ulster.ac.uk

Trudy Goodenough; e-mail: trudy.goodenough@bristol.ac.uk

Elizabeth Bailey; e-mail: elizabeth.bailey@uhcw.nhs.uk

Jane Coad; e-mail: jane.coad@nottingham.ac.uk

Crispin Day; e-mail: Crispin.1.day@kcl.ac.uk

Samantha Nightingale; e-mail: samantha.nightingale@uhcw.nhs.uk

Sally Kendall; e-mail: S.Kendall-608@kent.ac.uk

Raghu Lingam; e-mail: r.lingam@unsw.edu.au

*Contributions* (I) Conception and design: T Deave, R Lingam, J Coad, S Kendall, C Day; (II) Administrative support: T Goodenough, S Ginja; (III) Provision of study materials or patients: T Deave, T Goodenough, S Ginja, E Bailey, S Nightingale; (IV) Collection and assembly of data: T Goodenough, S Ginja, E Bailey, S Nightingale; (V) Data analysis and interpretation: S Ginja, T Deave, E Bailey, S Nightingale, J Coad, R Lingam, C Day, S Kendall; (VI) Manuscript writing: All authors; (VII) Final approval of manuscript: All authors.

**Joint first authors*

¥*Corresponding Author*: Toity Deave; e-mail: Toity.deave@uwe.ac.uk

Dr Toity Deave iHV Fellow, Assoc. Professor for Family and Child Health Centre for Academic Child Health, University of the West of England, Bristol, 1-5 Whiteladies Road, Bristol BS8 1NU, Tel: +44 117 428 3094

Short running title: Baby Buddy app evaluation: cohort study

4835 word count

## Abstract

Background:

Health mobile applications (apps) have become very popular, including apps specifically designed to support women during the ante- and postnatal periods. However, there is currently limited evidence for the effectiveness of such apps at improving pregnancy and parenting outcomes.

Aim: to assess the effectiveness of a pregnancy and perinatal app, Baby Buddy, in improving maternal self-efficacy at three months post-delivery.

Methods: Participants were 16-years and over, first-time pregnant women, 12-16 weeks gestation, recruited from five English study sites. The Tool of Parenting Self-efficacy (TOPSE) (primary outcome) was used to compare mothers at three months post-delivery who had downloaded the Baby Buddy app compared to those who had not downloaded the app, controlling for confounding factors.

Results: 488 participants provided valid data at baseline (12-16 weeks gestation), 296 participants provided valid data at 3 months post-birth, 114 (38.5%) of whom reported that they had used the Baby Buddy app. Baby Buddy app users were more likely to use pregnancy or parenting apps (80.7% vs 69.6%, p=.035), more likely to have been introduced to the app by a healthcare professional (p=.005) and have a lower median score for perceived social support (81 vs 83, p=.034) than non-app users. The Baby Buddy app did not illicit a statistically significant change in TOPSE scores from baseline to 3 months post-birth (adjusted OR 1.12, 95%CI 0.59 to 2.13, p=.730). Finding out about the Baby Buddy app from a healthcare professional appeared to grant no additional benefit to app users compared to all other participants in terms of self-efficacy at three

months post-birth (adjusted OR 1.16, 95%CI 0.60 to 2.23, p=.666). There were no statistically significant differences in the TOPSE scores for the in-app data between either the type of user who was engaged with the app and non-app users (adjusted OR 0.69, 95%CI 0.22 to 2.16, p=.519) or those who were highly engaged and non-app users (adjusted OR 0.48, 95%CI 0.14t o 1.68, p=.251).

Conclusion: This study is one of few, to date, that has investigated the effectiveness of a pregnancy and early parenthood app. No evidence for the effectiveness of the Baby Buddy app was found. New technologies can enhance traditional healthcare services and empower users to take more control over their healthcare but app effectiveness needs to be assessed. Further work is needed to consider, a) how we can best use this new technology to deliver better health outcomes for health service users and, b) methodological issues of evaluating digital health interventions.

## Keywords

Evaluation, first-time parents, Baby Buddy, self-efficacy, maternal well-being.

# The Bumps and BaBies Longitudinal Study (BaBBLeS): a multi-site cohort study of first-time mothers to evaluate the effectiveness of the Baby Buddy app

## Introduction

Electronic (e-Health) and mobile (m-Health) health methodologies are increasingly used to improve the self-management of health problems in many countries (1). This change in health seeking behaviour has been influenced by easier internet access, greater device functionality and poorer access to face-to-face healthcare services. There has been a growing interest in the capability of smartphone applications ('apps') to promote health, encourage behaviour change and enhance the service users' experience. There are over 318,000 health apps currently available on the leading app stores, with more than 200 apps added daily (2). However, systematic reviews have demonstrated that evidence of the effectiveness of health behaviour change apps remains limited and that studies of better quality are needed (3-5).

Ante- and post-natal care are one of the domains that has seen a large expansion of mobile apps. There are thousands of apps focused on women's health and pregnancy, corresponding, approximately, to 7% of all existing health apps (6). It is commonly assumed that such apps have the potential to enhance conventional pregnancy and postnatal care (7). However, consistent with the wider literature on health apps, two systematic reviews found limited evidence of the effectiveness of apps designed specifically for ante- and/or post-natal care or women's health (8,9). Although these

reviews found a small number of evaluation studies where an experimental design had been used, they stressed the need for more high quality studies and with adequately powered samples, as well as the need to assess the validity of app contents. It was also reported that, whilst some pregnancy and parenting app types have been assessed in a number of studies (e.g., gestational weight gain prevention), others, such as mental health-related apps, are lacking (9). The Baby Buddy app was developed by the national child health and wellbeing charity, 'Best Beginnings'. Its public health purpose was to provide evidence-based, professionally validated information to pregnant and new mothers, empower women's positive pregnancy and early parenting health behaviours, promote contacts with healthcare professionals and increase mothers' self-efficacy with regard to pregnancy, babycare and early parenthood (10). Parental well-being and self-efficacy, that is parents' self-perception about their ability to perform as parents, are major determinants of child health and development, parent-child relationships and buffer against parenting stress(11–13). The app content and functionality was co-created with parents and professionals and had a minimum reading age of 11 years with a 'read aloud' element available. It included interactive information to help parents manage their physical and mental health and to help them to support the physical and emotional health of their child. It was designed to complement maternity and postnatal services and support the aim of 'making every contact count'(14). Integration with health service delivery was promoted by Best Beginnings on the basis that mothers introduced to the app by a healthcare professional maybe more likely to use it.

Based on 'proportionate universalism'(15), Baby Buddy was intended to be used by mothers across the age-range with a particular focus on engaging groups at higher risk

of poorer outcomes, such as expectant mothers under 25-years old. These younger mothers are less likely to engage with maternity services early in pregnancy and less likely to attend maternity appointments (16). Both behaviours are risk factors for adverse pregnancy outcomes (17). Baby Buddy was available for download by expectant mothers, partners, family members and friends from Apple iStore and the Google Play. Download data recorded by the app developers appeared to support its use by younger mothers(10).

The aim of the Bumps and BaBies Longitudinal Study (BaBBLeS) reported in this paper was to assess the effectiveness of the Baby Buddy app on improving maternal self-efficacy and mental wellbeing.

## Methods

This longitudinal, mixed methods study was conducted in five geographical sites in England. It had three component parts: a cohort study, analysis of in-app data and a qualitative study. The study protocol has been previously published (18). An Appreciative Approach was used for the qualitative study with the results published elsewhere (19). This paper reports on findings from the cohort study and in-app data analysis.

The cohort study compared self-reported self-efficacy and mental wellbeing of (i) mothers three months post-delivery who had used the Baby Buddy app with those mothers who had not, and (ii) mothers who were shown how to use the app by a health professional, as advocated by the app developers, compared to those who were not

shown or did not download it. In-app data were collected on uptake, usage pattern and detailed analytics of key app functionality.

Recruitment took place between September 2016 and February 2017. Women aged 16 years and over, with no previous live child, and between 12-16 weeks and six days gestation were identified by the participating maternity units in the five study sites. Each identified woman was sent or given a study invitation letter and information booklet. Mothers completed questionnaires, online or on paper, which comprised of quantitative outcome measures and sociodemographic questions. A £5 voucher was issued upon receipt of the completed questionnaire (appendix 1). A two week reminder was sent if no questionnaire was received.

**Data collection**

Cohort study

Quantitative data were collected at three time points: 12-16 weeks pregnancy (baseline), 35 weeks pregnancy and 3 months post-birth. This paper focusses on the data collected at baseline and at three months' post-birth. The 35 weeks gestation data did not affect these results. All data were obtained from participant self-report.

At baseline, women provided informed consent for cohort study participation and completed the required measures.

In-app data

At the 35-week gestation data collection, mothers were sent an information sheet and consent form to complete in order to take part in this element of the study. The majority of Baby Buddy app use patterns were recorded and stored on secured databases, hosted by Best Beginnings, as part of a standard procedure necessary for managing and

159 debugging the app. For those mothers who gave their consent, using anonymised
160 personal identification codes, Best Beginnings provided the research team with limited
161 and secured download access to the database to obtain specific in-app data from app
162 users, including duration of app use sessions, app session count, app use flow, and
163 general user information.

164 **Outcome measures**

165 1. Primary outcome

166 Tool to measure Parenting Self-Efficacy (TOPSE) (20,21).

167 The primary cohort study outcome measure was the TOPSE which is underpinned by
168 self-efficacy theory (22). The TOPSE shorter version is a multi-dimensional instrument
169 of 36 items within six scales representing distinct dimensions of parenting: emotion
170 and affection, play and enjoyment, empathy and understanding, pressures, self-
171 acceptance, learning and knowledge. The items are rated on an 11-point Likert scale,
172 0 (completely disagree) to 10 (completely agree), responses are summed to create a
173 total score, lower scores indicating lower parenting self-efficacy. Subscale internal
174 reliability coefficients ranged 0.80 to 0.89 and overall scale reliability was 0.94.
175 External reliability coefficients ranged from rs = 0.58 (n=19, $p<0.01$) to rs = 0.88 (n=19,
176 $p<0.01$). The 0-6 month version of TOPSE was adapted, in collaboration with the
177 author, to measure parenting self-efficacy expectations during pregnancy.

178 2. Secondary outcome

179 Warwick-Edinburgh Mental Well-Being Scale (WEMWBS) (23).

180 The WEMWBS was the secondary outcome measure validated for use in the UK with
181 those aged 16 and above. It is a 14 item scale of subjective mental well-being and

psychological functioning describing feelings (eg.,'I have been feeling useful') and functional aspects (eg., 'I've been dealing with problems well') over the previous two week. Items are scored from 1 (none of the time) to 5 (all of the time) and summed to provide an overall score between 14 and 70, where higher scores corresponded to greater frequency. WEMWBS has good content and criterion-related validity and high test-retest reliability (0.83,(24)).

**Sociodemographic variables**

Sociodemographic and health data collected included women's age, ethnic group, socio-economic deprivation, highest level of formal education, relationship status and employment. Index of multiple deprivation (IMD) decile, a common indicator of socioeconomic deprivation in the UK, was obtained by searching participants' postcodes using a standard online tool (25). The geographical site where participants were recruited was also noted. Social support was measured using the Multidimensional Scale of Perceived Social Support (MSPSS (26)) and technology use was assessed using the Media and Technology Usage and Attitudes Scale (MTUAS) (27). In addition, at baseline and at 35 weeks gestation, participants' expected date of delivery (EDD) and intended baby feeding methods was recorded. At three months post-birth, information about participants' childbirth experience, using the Childbirth Experience Questionnaire (CEQ) (28), and actual baby feeding methods was collected. For more details see the published protocol (29).

**Sample size**

Our original sample size calculation assumed linearity of outcome variables (18). Both primary and secondary outcomes were negatively skewed and therefore converted to

dichotomous variables, lowest quartile compared to the upper three quartiles. The original sample size of 559 women assumed a 12.5% app download, which meant roughly a ratio of 1 Baby Buddy user to 7 non-users (29). However, as explained in the results section, the percent app download was higher than anticipated which reduced the required sample size to 250 participants (due to a smaller ratio). This included 100 intervention subjects (i.e. Baby Buddy app users) and 150 controls (i.e. non-app users) to have 80% power to detect a 7% difference (0.5 SD) in the proportion of participants in the lowest quartile compared to the upper three quartiles at the 5% level (30).

**Data analysis**

Descriptive statistics were used to describe the sample, including the mothers' age, socio-demographics, ethnicity, access to and use of technology and the overall sum scores for the outcome measures. Logistic regression models were used to compare the primary and secondary outcomes in mothers who used the Baby Buddy app compared to those who did not use the app. Participants were considered app users if they had reported using the app at any of the three data collection time points. Logistic regression diagnostics using Hosmer and Lemeshow's goodness-of-fit test indicated a good fit of the adjusted models ($p>.05$). Key variables were tested as potential confounders, including maternal age, education, employment, relationship status, recruitment site, social support, general technology use and use of other pregnancy apps. Baseline levels of the outcome variables were also controlled for in the final analysis. Analysis was as per protocol and analysis plan unless otherwise specified. All analyses were carried out using Stata 14 software.

The TOPSE scores were negatively skewed so a log transformation of these data was carried out but the distribution remained non-normal. As a result, we developed logistic regression models in which TOPSE scores were converted into a binary variable: low self-efficacy (1), to represent those in the lowest quartile of TOPSE score data and reference levels of self-efficacy (0), which corresponded with those with TOPSE scores above the lowest quartile. In this analysis, we report the odds ratio of low TOPSE scores (i.e., low self-efficacy) amongst Baby Buddy app users compared to non-app users. This logistic regression analysis comprised of two models: i) unadjusted model and, ii) model adjusted for potential confounders, including baseline levels of the outcome.

Secondary analysis

A second analysis compared primary and secondary outcomes, as described above, between those mothers who used the app and heard about it from a health professional (instructed use) and those women who did not hear about it or who did not download the app by three months post-delivery.

Post-hoc analysis

Qualitative findings suggested that Baby Buddy breastfeeding contents were popular (19). It was decided to conduct a post-hoc analysis of the impact of the Baby Buddy app on self-reported breastfeeding.

In-app data:

For consenting mothers (n=51), uptake, patterns of usage and detailed analytics of key factors within the app were analysed. These were participants who had provided valid

outcome data at baseline (i.e., TOPSE or WEMWBS data) and who also responded at three months post-birth with valid outcome data.

Data orientation was undertaken and then formatted for analysis. This included an exploratory analysis of socio-demographic information and profiling of app users (e.g. age, occupation, education, ethnic origin); description of app use patterns including the creation of the app avatar; goal setting function, media downloaded and the app functions of 'ask me a question' and 'what does that mean.

In consultation with the app developers, the following app elements were assessed to quantify in-app usage: 'Today's Information', 'Videos', 'Ask Me', 'Remember to Ask', 'You can Do it', 'Bump Around/Baby Around', 'Baby Book/Bump Book', 'Baby Booth/Bump Booth', and 'What Does it Mean'. Further details of these app functions are provided in the appendix. The number of times each element of the app was used were summed and two overall aggregated scores were derived for data analysis. The first score was a 'passive' overall score, based exclusively on the 'Today's Information' element. This included whether this feature had been opened, if links were followed and whether participants tapped on 'Read more'. This involved mostly viewing and clicking information and was less goal- and behaviour change-oriented. The second composite score was an 'active' overall score and encompassed all other app elements. This was a more proactive format of app interaction, for example, users had to specifically search for information or videos or set up reminders.

Based on the median value of the session count, the passive users were sub-divided into passive high users (n=26; 94 sessions or more) and passive low app users (n=25; less than 94 sessions). Similarly, the active high app users (n=27; 27 sessions or more) and

active low app users (n=24; less than 27 sessions) sub-divided into two groups. Separate logistic regression models were developed to compare outcomes (TOPSE and WEMWBS, as dichotomised in previous models) between active high and low app users and passive high and low app users. The same two regression models used for the questionnaire data were performed, one unadjusted (model 1) and one adjusted for potential confounders (model 2). However, considering the small number of participants in the analyses, to maximise the viability of the model, there had to be careful selection of the confounding variables to be included. Differences between high/low app users were analysed and confounding factors were selected which were shown to be significant at the baseline outcome level for TOPSE and WEMWBS.

### Ethics

This study received a favourable opinion from the NHS Research Ethics Committee (NRES) West Midlands-South Birmingham REC (16/WM/0029), the University of the West of England, Bristol Research Ethics Committee (HAS.16).

## Results

### Descriptive results

A total of 488 participants provided valid data at baseline, i.e., TOPSE data and/or WEMWBS data (initial sample). Of this initial sample, 256 participants (52.5%) provided valid data at 35 weeks gestation. Of the initial sample, 296 (60.7%) provided valid data at 3 months post-birth; this was the sample used in the main analysis, hereinafter referred to as the final sample. There were 220 participants (45.1%) who provided data at all three data collection time-points. The participant flow is presented in figure 1.

Of the 296 participants followed to 3 months post-birth, 114 reported to be Baby Buddy app users (38.5%), i.e. they had reported using the Baby Buddy app at one or more of the three data collection time-points. This corresponds roughly to a ratio of 1 to 2, i.e. one reported Baby Buddy user for every two non- Baby Buddy users.

The distribution of participants in the initial sample (N=488) by recruitment site was as follows: 168 from the West Midlands (34.4%), 139 from London (28.5%), 66 from West Yorkshire (13.5%), 62 from Lancashire (12.7%) and 53 from East Midlands (10.9%). This distribution, per site, remained very similar in the final sample. Baseline characteristics of participants included in the final sample are presented by app use in table 1. App users (n=114) were comparable to non-app users (n=182) in age, Index of Multiple Deprivation (IMD) decile, ethnicity, highest education attained, employment and relationship status.

All participants used a mobile phone and had internet access and nearly all had internet at home. Two thirds used a tablet. There were no significant baseline differences between Baby Buddy users and non- Baby Buddy users in terms of any of these variables. The three top sources of information about pregnancy and parenthood, in both groups, were the internet (app users 88.5%; non-app users 82.7%), friends (app users 82.4%; non-app users 76.5%) and midwife (app users 74.3%; non-app users 71.0%). For both Baby Buddy users and non- Baby Buddy users, the overall median MTUAS score was 5. No significant differences with regards to any of these variables were observed between the two groups. There are no set thresholds to distinguish between ‘high technology use’ and ‘low technology use’, so comparison between group scores were made(31).

Baby Buddy users were significantly more likely to use pregnancy/parenthood apps in general, not just the Baby Buddy app, than non- Baby Buddy users at baseline (80.7% vs 69.6%, p=.035) consequently, this was one of the variables adjusted for in the main analysis. Baby Buddy users were also more likely to have heard about the pregnancy apps they used from healthcare professionals than non- Baby Buddy users (p=.005). On the overall MSPSS score, Baby Buddy users had a significantly lower median score (81) than non- Baby Buddy users (83), p=.034; this indicates lower levels of perceived social support amongst Baby Buddy users at baseline.

Baseline data for the outcome variables show that the median score for the TOPSE was 317 (287-337, LQ-UQ) for app users 320 (295-337, LQ-UQ) for non-app users (table 2). For the WEMWBS, the median for app users and non-app users were 54 (49-59, LQ-UQ) and 54 (48-61, LQ-UQ), respectively. There were no statistically significant differences between the two groups for either the TOPSE or WEMWBS. Similar to the MSPSS, TOPSE and WEMWBS scores are used for comparison between participants or across time.

**Outcome results**

At 3 months post-birth, there were no statistically significant differences in TOPSE or and WEMWBS outcomes between Baby Buddy users and non- Baby Buddy users. Baby Buddy users had a median TOPSE score of 319 (LQ 296 – UQ 338) compared to non-Baby Buddy users who had a median TOPSE score of 327 (LQ 305 – UQ 343), p=.107. Similarly, Baby Buddy users had a median WEMWBS score of 54.5 (LQ 49 – UQ 59) compared to non- Baby Buddy users who had a median score of 55 (LQ 50 – UQ 61), p=.284.

The unadjusted odds ratio for low TOPSE score (i.e. lower self-efficacy) was 1.17 (95% CI 0.68 to 2.03, p=.564) amongst Baby Buddy users compared to non-Baby Buddy users (table 3). Adjustment of this association for IMD decile, technology use (baseline MTUAS total mean score), use of pregnancy/parenthood apps (any), social support (baseline MSPSS overall sum score) and baseline TOPSE score resulted in a very similar result: adjusted odds ratio of 1.12 (95%CI 0.59 to 2.13, p=.730). The Baby Buddy app had no significant effect on maternal mental wellbeing, with an unadjusted odds ratio for low WEMWBS of 1.10 (95% CI 0.64 to 1.89, p=.719). Adjustment for confounding factors made minimal difference to this association, OR 1.02 (95% CI 0.55 to 1.89, p=.943)(table 3).

Baby Buddy users who had heard about the app from a healthcare professional had slightly higher odds of a low self-efficacy TOPSE scores compared to all other participants. These differences were not statistically significant, neither in the unadjusted model (model 1) (OR 1.16, 95%CI 0.66 to 2.04, p=.596) nor in the adjusted model (model 2) (OR 1.16, 95%CI 0.60 to 2.23, p=.666). Similarly, there were no differences in the odds ratios for low WEMWBS scores between Baby Buddy users who had heard about the app from a healthcare professional and all other participants, neither in the unadjusted model (OR 1.03, 95%CI 0.59 to 1.79, p=.924) nor in the adjusted model (OR 1.00, 95%CI 0.53 to 1.87, p=.990).

**In-app data**

The number of uses of each aggregated score: passive, active and the overall usage, see table 4, suggest that participants engaged more with the passive elements of the app.

Changes in levels of app usage and whether they affected the reported outcomes (i.e. TOPSE and WEMWBS scores) were explored. The differences between the characteristics of in-app participants (those who had consented to their in-app data being used and who had provided valid outcome data at baseline and 3 months post-birth (n=51) and non- Baby Buddy users (n=182) were similar to those differences between Baby Buddy users and non- Baby Buddy users, i.e., statistically non-significant except that in-app users had lower social support (p=.035) and used more pregnancy/parenthood apps than non- Baby Buddy users (p<.0001).

The results of the logistic regression analysis for both self-efficacy (TOPSE) and mental wellbeing (WEMWBS) and any association with usage of the passive and active in-app elements are described in table 5. For clarity, we also report the median value of the outcome score, for each of the two groups (under the columns 'High users' and 'Low users'). The results revealed no statistically significant associations between level of usage of the passive in-app element and TOPSE scores, and WEMWBS scores, neither in the unadjusted nor in the adjusted models. Confidence intervals were large, particularly for WEMWBS. Another set of analyses were performed comparing high app users with non- Baby Buddy users, rather than with low users. Results, not reported here, were very similar to those presented in table 5, with no statistically significant differences between the two groups.

<u>Post-hoc analysis on breastfeeding</u>

Baby Buddy users were more likely to report that they had breastfed at 1 week post-birth, at 1 month post-birth and at 3 months post-birth (table 6). This included breastfeeding in combination with formula milk ('any breastfeeding') and breastfeeding

as the sole baby feeding method ('exclusive breastfeeding'). At 1 month post-birth, this difference was statistically significant for both any breastfeeding, (X2 (1) = 10.68, p=.001) and exclusive breastfeeding (X2 (1) = 3.86, p=.05) (table 6).

Logistic regression models were developed to explore the association between breastfeeding and Baby Buddy use, using the same unadjusted and adjusted models from the main analysis (table 7). At all time-points, Baby Buddy app users had increased odds of breastfeeding compared to non- Baby Buddy users. However, differences between the two groups were only statistically significant for any breastfeeding at 1 month post-birth, both unadjusted (OR 2.68, 95%CI 1.46 to 4.90, p=.001) and after adjusting for confounding variables (OR 3.08, 95%CI 1.49 to 6.35, p=.002) and at 3 months post-birth in the adjusted model for exclusive breastfeeding (OR 1.79, 95%CI 1.02 to 3.16, p=.044)(table 7).

## Discussion

There is a lack of evidence about the effectiveness of pregnancy/parenthood apps with those studies that aim to assess this being insufficiently powered to detect significant effects (8,9). The BaBBLeS study aimed to address this research gap by being one of the first large-scale controlled studies to assess the effectiveness of such an app, Baby Buddy, at improving reported maternal psychological outcomes. Our findings suggested that the app had no effect on maternal parenting self-efficacy and mental wellbeing at three months post-birth. There were also no statistically significant outcome differences between those who used the app more than the median number of app sessions and those who used it less, based on objective (in-app) data, or between those who were

404 told about the app by a healthcare professional and those who found out about it
405 through other sources.

406 Although the use of the Baby Buddy app did not impact on the pre-specified outcomes,
407 a post-hoc analysis suggested that it did lead to higher levels of self-reported
408 breastfeeding, after adjusting for baseline differences and other relevant confounders.
409 These findings, though preliminary, are hypothesis generating and potentially
410 encouraging. Nevertheless, as a post-hoc analysis the findings require further
411 exploration using a pre-specified plan of analysis, ideally in a randomised controlled trial.
412 This is particularly important given its relevance to the current public health agenda. The
413 exploration of which specific features of the app are responsible for the improvements
414 in breastfeeding would be helpful for healthcare practitioners, especially midwives and
415 health visitors, so that those features could be emphasised in their contact with
416 mothers.
417 Midwives were the most frequent source of information about Baby Buddy, suggesting
418 that the app developers were successful in their maternity dissemination methods with
419 the aim to 'make every contact count' (32). However, findings suggested that the app
420 may not lead to the expected improvements in maternal self-efficacy and mental well-
421 being even when integrated into in service delivery. However, improvements in non-
422 hypothesised outcomes such as breastfeeding were detected.
423 The lack of expected outcome impact may be due to the absence of the interpersonal
424 and personalised aspects of care that are core elements of face-to-face clinical
425 interactions (e.g., 33,34). It may be that apps may have a supplementary role but are

unlikely to replace direct clinical care especially when managing the challenges affecting the lives of vulnerable women during pregnancy and early infancy (35,36).

**Strengths and limitations of the study**

Outcome data were based on self-report using well-validated scales used previously to detect significant increases in self-efficacy and mental wellbeing. The TOPSE was adapted for antenatal use and the effect of anticipated, compared to actual, self-efficacy, on post-birth optimism is unknown. Outcome scores on both TOPSE and WEMWBS were high at baseline in app user group and the non-app user groups, raising the potential of ceiling effects. There was little change in total scores at each time point, inferring that the participant cohort was generally high functioning in parenting self-efficacy and mental wellbeing. While the app may have sought to influence these outcomes, participants expressed preference for talking to healthcare professionals face-to-face and to be with other parents (19).

The study used a broad definition of 'Baby Buddy user' that included any use of the app during the study period. This definition is consistent with an intention to treat approach but may lack sensitivity to the use of specific app functionality. The secondary analysis using the in-app data, however found no differences between high and low/no app users. This suggests that the lack of association between outcomes and Baby Buddy use was unlikely to have been due to measurement errors.

A longer, e.g., six-month, follow up period may have been preferable. However a systematic review of web-based interventions for perinatal mood disorders suggests that three-month follow-up assessments can detect outcome improvement (37).

Using a randomised, rather than quasi-experimental, design would strengthen the inferences drawn from the study's findings. However, randomisation was not possible because the Baby Buddy app was freely available for download, risking contamination in those randomised to a comparison condition. Furthermore, the only difference between Baby Buddy app using and non-app using mothers at baseline was the use of other maternity apps by the Baby Buddy app-using mothers, which suggests that mothers may either be users of several apps or none (38).

We are unable to provide an estimate of the proportion of women approached by midwives who agreed to study participation. While using recruitment logs, maternity staff limitations, prevented them from being anonymised and then shared with the research team. Retention rates in studies involving ante- and post-natal women are variable but the study's 60% rate is consistent with those reported in clinical research trials involving perinatal women (39,40). It attests to the difficulty of engaging with new mothers at such a demanding period of their lives. The final sample included just those mothers who had complete data for the TOPSE and WEMWBS at baseline and at three months post-birth. The baseline characteristics of those mothers in the final sample largely reflected those of the initial sample and app users and non-app users remained comparable.

Participants were self-selected and we were unable to assess their representativeness for the wider population of first-time mothers in each site. The sample was predominantly composed of White British women living in areas of higher economic deprivation (41). However, the rate of degree holders, at baseline, 51.0% and in the final sample, 58.6%, is substantially higher than the national average of 42% (42). This was

affected by the characteristics of the London site, where a considerable part of our sample was based. The greater likelihood of more socially advantaged participants is a common phenomenon in maternal health-related research(43,44).

## Conclusions

There is an increasing emphasis on the use of technologies to support the delivery of healthcare services, as evident from the National Health Service apps library (45). New technologies may have potential to enhance and even replace conventional healthcare provision as well as empower people to take more control over their healthcare. This is one of the few studies to date to investigate the health outcomes of a specific app designed for use by mothers in the antenatal and early postnatal periods. It found no evidence of impact on first-time mothers' self-reported parental self-efficacy and mental well-being at three months post-birth though post-hoc analysis suggested that app users were more likely to exclusively breastfeed, or ever breastfeed. Overall findings suggest that this particular app may have limited impact on the outcomes measured. Further work is needed to differentiate the types of outcomes the app may improve as well as how new technologies more widely can best optimise to health outcomes.

## List of abbreviations

IMD: Index of Multiple Deprivation

MSPSS: Multidimensional Scale of Perceived Social Support

MTUAS: Media and Technology Usage and Attitudes Scale

493 NHS: National Health Service

494 TOPSE: Tool of Parenting Self-efficacy

495 WEMWBS: Warwick and Edinburgh Mental Well-being Scale

## 496 Acknowledgements

497 The authors would like to thank all the participants of this study – the mothers and the
498 health professionals. They would also like to thank the five participating midwifery
499 services who supported and undertook the process of recruitment to the study and
500 follow-up data collection.

501

502 Funding: This work was supported by the Big Lottery via Best Beginnings as a competitive
503 tender.

504

## 505 Disclosure

506 The authors have no conflict of interest, neither financial nor personal.

507

## 508 Footnote

509 The authors are accountable for all aspects of the work in ensuring that questions
510 related to the accuracy or integrity of any part of the work are appropriately investigated
511 and resolved.

512

## 513 References

1. Zapata BC, Fernández-Alemán JL, Idri A *et al.* Empirical Studies on Usability of mHealth Apps: A Systematic Literature Review. J Med Syst. 2015 Feb 20;39(2):1.

2. IQVIA. The Growing Value of Digital Health [Internet]. 2017 [cited 2019 Mar 13]. Available from: https://www.iqvia.com/institute/reports/the-growing-value-of-digital-health

3. Zhao, J., Freeman, B. and Li, M. (2016) 'Can Mobile Phone Apps Influence People's Health Behavior Change? An Evidence Review', *Journal of Medical Internet Research*, 18(11), p. e287. doi: 10.2196/jmir.5692.

4. Marcolino, M. S. *et al.* (2018) 'The Impact of mHealth Interventions: Systematic Review of Systematic Reviews', *JMIR mHealth and uHealth*, 6(1), p. e23. doi: 10.2196/mhealth.8873.

5. McKay FH, Cheng C, Wright A, *et al.* Evaluating mobile phone applications for health behaviour change: A systematic review. J Telemed Telecare. 2018 Jan 18;24(1):22–30.

6. Aitken, M., & Lyle, J. (2015). Patient adoption of mHealth: use, evidence and remaining barriers to mainstream acceptance. Parsippany, NJ: IMS Institute for Healthcare Informatics. Available at: https://www.iqvia.com/-/media/iqvia/pdfs/institute-reports/patient-adoption-of-mhealth.pdf (Accessed: 5 July 2019).

7. Tripp N, Hainey K, Liu A, *et al.* An emerging model of maternity care: Smartphone, midwife, doctor? Women Birth. 2014 Mar 1;27(1):64–7.

8. Overdijkink, S. B. *et al.* (2018) 'The Usability and Effectiveness of Mobile Health Technology–Based Lifestyle and Medical Intervention Apps Supporting Health Care

During Pregnancy: Systematic Review', *JMIR mHealth and uHealth*, 6(4), p. e109. doi: 10.2196/mhealth.8834.

9. Derbyshire, E. and Dancey, D. (2013b) 'Smartphone Medical Applications for Women's Health: What Is the Evidence-Base and Feedback?', *International Journal of Telemedicine and Applications*, 2013, pp. 1–10. doi: 10.1155/2013/782074.

10. Best Beginnings. Best Beginnings. About Baby Buddy [Internet]. 2017 [cited 2018 Sep 17]. Available from: https://www.bestbeginnings.org.uk/about-baby-buddy

11. Coleman PK, Karraker KH. Maternal self-efficacy beliefs, competence in parenting, and toddlers' behavior and developmental status. Infant Ment Health J. 2003;24(2):126–48.

12. Deave T, Heron J, Evans J, *et al*. The impact of maternal depression in pregnancy on early child development. BJOG Int J Obstet Gynaecol. 2008;115(8):1043–51.

13. Kendall S, Bloomfield L. Developing and validating a tool to measure parenting self-efficacy. J Adv Nurs. 2005;51(2):174–81.

14. Hemmings P. MECC Consensus statement. Consens Statement. :18.

15. Marmot MG, Allen J, Goldblatt P, *et al*. Fair society, healthy lives: Strategic review of health inequalities in England post-2010 [Internet]. London UK: The Marmot Review; 2010 Feb [cited 2019 Apr 30]. Available from: http://discovery.ucl.ac.uk/111743/

16. Bradshaw P, Schofield L, Maynard L. Health and Social Care The Experiences of Mothers Aged Under 20: Analysis of Data From the Growing Up in Scotland Study. 2014.

17. Raatikainen K, Heiskanen N, Heinonen S. Under-attending free antenatal care is associated with adverse pregnancy outcomes. BMC Public Health. 2007 Dec 27;7(1):268.

18. Deave T, Kendal S, Lingam R, *et al.* A study to evaluate the effectiveness of Best Beginnings' Baby Buddy phone app in England: a protocol paper. Prim Health Care Res Dev. 2018 Jul 23;1–6.

19. Deave T, Coad J, Day C, *et al*. Bumps and Babies Longitudinal Study (BABBLES): An independent evaluation of the Baby Buddy app [Internet]. 2018 [cited 2019 Apr 30]. Available from: http://eprints.uwe.ac.uk/39012/

20. Kendall S, Bloomfield L. Developing and validating a tool to measure parenting self-efficacy. J Adv Nurs. 2005 Jul;51(2):174–81.

21. Bloomfield L, Kendall S. Parenting self-efficacy, parenting stress and child behaviour before and after a parenting programme. Prim Health Care Res Dev. 2012 Oct 2;13(04):364–72.

22. Bandura A. Self-efficacy mechanism in human agency. Am Psychol. 1982;37(2):122–47.

23. Tennant R, Hiller L, Fishwick R, *et al.* The Warwick-Edinburgh Mental Well-being Scale (WEMWBS): development and UK validation. Health Qual Life Outcomes. 2007 Nov 27;5(1):63.

24. Stewart-Brown SL, Platt S, Tennant A, *et al.* The Warwick-Edinburgh Mental Well-being Scale (WEMWBS): a valid and reliable tool for measuring mental well-being in diverse populations and projects. J Epidemiol Community Health. 2011 Sep 1;65(Suppl 2):A38–9.

25. English indices of deprivation 2015 [Internet]. GOV.UK. [cited 2019 Apr 30]. Available from: https://www.gov.uk/government/statistics/english-indices-of-deprivation-2015

26. Zimet GD, Dahlem NW, Zimet SG, *et al*. The Multidimensional Scale of Perceived Social Support. J Pers Assess. 1988;52(1):30–41.

27. Rosen LD, Whaling K, Carrier LM, *et al.*. The Media and Technology Usage and Attitudes Scale: An empirical investigation. Comput Hum Behav. 2013 Nov 1;29(6):2501–11.

28. Dencker A, Taft C, Bergqvist L, *et al.* Childbirth experience questionnaire (CEQ): development and evaluation of a multidimensional instrument. BMC Pregnancy Childbirth. 2010 Dec 10;10:81.

29. Deave T, Kendal S, Lingam R, *et al*. A study to evaluate the effectiveness of Best Beginnings' Baby Buddy phone app in England: a protocol paper. Prim Health Care Res Dev [Internet]. 2019 ed [cited 2019 Apr 30];20. Available from: https://www.cambridge.org/core/journals/primary-health-care-research-and-development/article/study-to-evaluate-the-effectiveness-of-best-beginnings-baby-buddy-phone-app-in-england-a-protocol-paper/03A70C1641E3720179713C0001470EB6

30. Dupont WD, Plummer WD. Power and sample size calculations: A review and computer program. Control Clin Trials. 1990 Apr 1;11(2):116–28.

31. Rosen LD, Whaling K, Carrier LM, *et al.* The Media and Technology Usage and Attitudes Scale: An empirical investigation. Comput Hum Behav. 2013 Nov 1;29(6):2501–11.

605 32. Best Beginnings. Enhancing capacity of professionals & community [Internet].
606 2017 [cited 2019 Mar 13]. Available from:
607 https://www.bestbeginnings.org.uk/enhancing-capacity-of-
608 professionalscommunity

609 33. Seward MW, Simon D, Richardson M, *et al.* Supporting healthful lifestyles during
610 pregnancy: a health coach intervention pilot study. BMC Pregnancy Childbirth.
611 2018 Dec 17;18(1):375.

612 34. Willcox JC, van der Pligt P, Ball K, *et al.* Views of Women and Health Professionals
613 on mHealth Lifestyle Interventions in Pregnancy: A Qualitative Investigation. JMIR
614 MHealth UHealth. 2015 Oct 28;3(4):e99.

615 35. Santarossa S, Kane D, Senn CY, *et al.* Exploring the Role of In-Person Components
616 for Online Health Behavior Change Interventions: Can a Digital Person-to-Person
617 Component Suffice? J Med Internet Res. 2018 Apr 11;20(4):e144.

618 36. Prentice JL, Dobson KS. A review of the risks and benefits associated with mobile
619 phone applications for psychological interventions. Can Psychol Can.
620 2014;55(4):282–90.

621 37. Lee EW, Denison FC, Hor K, *et al.* Web-based interventions for prevention and
622 treatment of perinatal mood disorders: a systematic review. BMC Pregnancy
623 Childbirth. 2016 Dec 29;16(1):38.

624 38. Lupton D, Pedersen S. An Australian survey of women's use of pregnancy and
625 parenting apps. Women Birth. 2016 Aug 1;29(4):368–75.

626 39. Frew PM, Saint-Victor DS, Isaacs MB, *et al.* Recruitment and Retention of Pregnant
627 Women Into Clinical Research Trials: An Overview of Challenges, Facilitators, and
628 Best Practices. Clin Infect Dis. 2014 Dec 15;59(suppl 7):S400–7.

629 40. McCarter DE, Demidenko E, Hegel MT. Measuring outcomes of digital technology-
630 assisted nursing postpartum: A randomized controlled trial. J Adv Nurs. 2018 Sep
631 1;74(9):2207–17.

632 41. Ginja S, Coad J, Bailey E, *et al.* Associations between social support, mental
633 wellbeing, self-efficacy and technology use in first-time antenatal women: data
634 from the BaBBLeS cohort study. BMC Pregnancy Childbirth. 2018 Dec 12;18(1):441.

635 42. Office for National Services. *Graduates in the UK labour market - Office for*
636 *National Statistics*, *ONS.GOV.UK*. Available at:
637 https://www.ons.gov.uk/employmentandlabourmarket/peopleinwork/employme
638 ntandemployeetypes/articles/graduatesintheuklabourmarket/2017 (Accessed: 25
639 July 2019).

640 43. Braig S, Grabher F, Ntomchukwu C, *et al.* The Association of Hair Cortisol with Self-
641 Reported Chronic Psychosocial Stress and Symptoms of Anxiety and Depression in
642 Women Shortly after Delivery. Paediatr Perinat Epidemiol. 2016;30(2):97–104.

643 44. Feinberg ME, Jones DE, Roettger ME, *et al.* Preventive Effects on Birth Outcomes:
644 Buffering Impact of Maternal Stress, Depression, and Anxiety. Matern Child Health
645 J. 2016 Jan 1;20(1):56–65.

646 45. NHS. NHS Apps Library [Internet]. 2019 [cited 2019 Mar 13]. Available from:
647 https://www.nhs.uk/apps-library/

648