Early identification of postpartum depression using demographic, clinical, and digital phenotyping

Lisa Hahn[1,2], Simon B. Eickhoff[1,2], Ute Habel[3,4], Elmar Stickeler[5], Patricia Schnakenberg[1,3], Tamme W. Goecke[5,6], Susanne Stickel [3], Matthias Franz[7], Juergen Dukart [1,2] and Natalia Chechko[1,3]

Abstract

Postpartum depression (PPD) and adjustment disorder (AD) affect up to 25% of women after childbirth. However, there are no accurate screening tools for either disorder to identify at-risk mothers and enable them to benefit from early intervention. Combinations of anamnestic, clinical, and remote assessments were evaluated for an early and accurate identification of PPD and AD. Two cohorts of mothers giving birth were included in the study ($N = 308$ and $N = 193$). At baseline, participants underwent a detailed sociodemographic-anamnestic and clinical interview. Remote assessments were collected over 12 weeks comprising mood and stress levels as well as depression and attachment scores. At 12 weeks postpartum, an experienced clinician assigned the participants to three distinct groups: women with PPD, women with AD, and healthy controls (HC). Combinations of these assessments were assessed for an early an accurate detection of PPD and AD in the first cohort and, after pre-registration, validated in a prospective second cohort. Combinations of postnatal depression, attachment (for AD) and mood scores at week 3 achieved balanced accuracies of 93 and 79% for differentiation of PPD and AD from HC in the validation cohort and balanced accuracies of 87 and 91% in the first cohort. Differentiation between AD and PPD, with a balanced accuracy of 73% was possible at week 6 based on mood levels only with a balanced accuracy of 73% in the validation cohort and a balanced accuracy of 76% in the first cohort. Combinations of in clinic and remote self-assessments allow for early and accurate detection of PPD and AD as early as three weeks postpartum, enabling early intervention to the benefit of both mothers and children.

Introduction

The postpartum period poses the highest risk to women for developing a mental disorder[1], with postpartum depression (PPD) being the most frequent one[2]. PPD is defined as a major depressive disorder occurring in direct relation (within 4 weeks postpartum) to childbirth in the DSM-5[3]. Early diagnosis and treatment of PPD can substantially improve the outcome, prevent relapse, and minimize the associated emotional and financial burden[4].

Maternal mental health is a reliable predictor of child's cognitive development and subsequent achievements[5]. The risk of a mother-to-child transmission of the vulnerability to depression[6,7], through genetic as well as other factors such as depression-related effects on parenting[8], is particularly high. Successful treatment of maternal depression alleviates the risk of childhood behavioral problems[9].

PPD is often overlooked during postnatal visits, missing the critical window for early intervention[10,11]. One reason is that low mood in the early postpartum period is largely deemed "normal" with 50–80% of new mothers experiencing initial sadness (i.e., postpartum blues), primarily due to dramatically plunging hormone levels at parturition[12]. Adjustment disorder (AD) in reaction to postpartum stress is another postpartum condition with

Correspondence: Juergen Dukart (juergen.dukart@gmail.com) or
Natalia Chechko (nchechko@ukaachen.de)
[1]Institute of Neuroscience and Medicine, Brain & Behaviour (INM-7), Research Centre Jülich, Jülich, Germany
[2]Institute of Systems Neuroscience, Medical Faculty, Heinrich Heine University Düsseldorf, Düsseldorf, Germany
Full list of author information is available at the end of the article
These authors contributed equally: Juergen Dukart, Natalia Chechko

similar symptoms. The crucial difference to PPD is that the severity of AD does not meet the criteria for depression at any time point. In the clinical context, AD needs to be considered as an important differential diagnosis to PPD[13].

History of mental illness, vulnerability to hormonal changes, psychological and social distress, baby blues, premenstrual syndrome (PMS), unwanted pregnancy, traumatic birth experience and stressful life events are all associated with an increased risk of PPD[11,12,14]. It is of crucial importance to evaluate the relative and combined predictive value of these factors for development of PPD. Previous studies aiming at prediction of PPD focused either on time points in the late postpartum period (e.g., after 8–32 weeks)[15] or only on single time points, thereby ignoring symptom dynamics or convolving PPD with major depression or AD[16]. Detailed in-clinic assessments are costly and burdensome, providing the likely reason for the cross-sectional nature of most previous studies. Online remote self-assessments may provide an easy means of obtaining the relevant information on symptom dynamics in individual patients.

Here, we recruited two cohorts of mothers giving birth and followed them longitudinally over 12 weeks to explore whether an accurate prediction of PPD is feasible based on socio-demographic and clinical-anamnestic information as well as early symptom dynamics using remote mood and stress assessments. Data from the first cohort were used to identify combinations of demographic and clinical data achieving highest accuracy for early identification and differentiation of PPD and AD using a machine learning approach. In this cohort, we identified and trained the optimal model for individual diagnostic prediction. The model and approach were preregistered and evaluated against an independent validation cohort to obtain unbiased performance estimates of the proposed algorithm.

Methods
First cohort and study design
To identify the best predictors of PPD, a first cohort of 308 mothers (mean age = 31.7 ± 4.76) was recruited following childbirth at the University Hospital Aachen between November 2015 and June 2018. The current project was part of the Risk of Postpartum Depression (RiPoD) study conducted at the University Hospital Aachen. The main exclusion criteria were a depressive episode (according to a clinical interview) at the time of recruitment and specific child health conditions (for details see supplementary material). The recruitment was conducted at the Department of Gynecology and Obstetrics within the first two to five days postpartum. Out of a total recruitment pool of ~1000 births per year, 50–60% of women were contacted (30% were directly excluded based on some exclusion criteria due to close collaboration with the obstetrics department) of which 50% were willing to participate and met the inclusion criteria. Written informed consent was obtained from all participants. The authors assert that all procedures contributing to this work comply with the ethical standards of the relevant national and institutional committees on human experimentation and with the Helsinki Declaration of 1975, as revised in 2008. All procedures involving human subjects/patients were approved by the Institutional Review Board of the Medical Faculty of RWTH Aachen University (EK 208/15). The study design comprised follow-up for 12 weeks with evaluation at five time points each three weeks apart (T0-T4) (Fig. S1). Evaluations were conducted at the clinic for T0 and T4 and via remote online questionnaires for T1 to T3. All women were asked to complete mood and stress assessments (scale from one to ten, ten being high) online on a bi-daily basis. Remote assessments were sent via e-mail. If three consecutive assessments were missed, a reminder was sent via e-mail, which allowed for close monitoring of the participation.

A clinical interview was conducted at T0 to ascertain current conditions. At T4, an experienced psychiatrist conducted a second clinical interview for a final diagnosis. Based on this interview, participants were assigned into one of three groups: healthy controls (HC, $N = 247$, 80.2%) without any sign of depression during the whole observation period, and women meeting DSM-5 criteria for PPD ($N = 28$, 9.1%) or AD ($N = 33$, 10.7%)[3]. In case of a depression, the Hamilton Depression Rating Scale[17] was administered. Clinical interviews were based on the DSM-5[3].

An sociodemographic-anamnestic questionnaire was used to obtain additional information about personal and socioeconomic status, psychiatric history, current pregnancy, child, breastfeeding at T0, postpartum blues (T4), PMS[18] (T4), subjective quality of support at home (T4), and breastfeeding at T4 (Table 1, Table S1). The Stressful Life Events Screening Questionnaire[19] was collected to assess encounter with stressful life events (T0) (Table 1). The Edinburgh Postnatal Depression Scale (EPDS)[20] was collected at all time points (T0-T4). Maternal attachment was evaluated from T1 through T4 using the Maternal Postnatal Attachment Scale (MPAS)[21].

Second cohort
For the second cohort, further referred to as validation cohort, 193 mothers (mean age = 32.7 ± 4.78) were recruited between November 2018 and January 2020 following the same protocol and study design as for the first cohort (Fig. S1). The prevalence rates in the validation cohort were 76.2% for HC ($N = 147$), 8.29% for PPD ($N = 16$), and 15.5% for AD ($N = 30$).

Table 1 Sociodemographic and anamnestic data for the first and second cohort.

Sociodemographic/anamnestic variable	First cohort				Second cohort			
	HC	PPD	AD	Statistical test	HC	PPD	AD	Statistical test
Age (in years)	31.9 ± 4.61 N = 247	30.4 ± 5.51 N = 28	31.4 ± 5.09 N = 33	$X^2(2, N= 308) = 2.60$ $p = .27$	33.1 ± 4.31 N = 146	31.8 ± 7.04 N = 16	31.5 ± 5.42 N = 30	$X^2(2, N= 193) = 3.40$ $p = 0.18$
Education (years)	13.8 ± 2.89 N = 240	12.4 ± 2.85 N = 27	13.4 ± 4.67 N = 33	$X^2(2, N= 300) = 3.04$ $p = 0.22$	14.6 ± 3.22 N = 139	14.3 ± 2.96 N = 16	14.08 ± 2.56 N = 26	$X^2(2, N= 179) = 0.77$ $p = 0.68$
Personal psychiatric history (no/yes)	220/27	16/12	19/14	$X^2(2, N= 308) = 34.5$ $p < 0.001$ [b,1,2]	118/26[c]	6/10	15/15	$X^2(2, N= 190) = 24.2$ $p < 0.001$ [b,1,2]
Familial psychiatric history (no/yes)	194/53	16/12	18/15	$X^2(2, N= 308) = 13.3$ $p = 0.001$ [b,1,2]	112/33	8/8	65/14	$X^2(2, N= 193) = 10.8$ $p = 0.005$ [b,2]
Birth complications (no/yes)	209/37	20/8	24/9	$X^2(2, N= 307) = 5.57$ $p = 0.062$	121/26	10/6	25/5	$X^2(2, N= 193) = 3.80$ $p = 0.15$
Subjective birth-related psychological traumas (no/yes)	215/29	19/9	20/13	$X^2(2, N= 305) = 21.1$ $p < 0.001$ [b,1,2]	124/14	12/4	23/7	[a]
PMS (no PMS/mild PMS/PMS)	111/84/29	4/12/12	7/16/10	$X^2(4, N= 285) = 27.9$ $p < 0.001$ [b,1]	83/44/14	3/4/8	9/14/6	$X^2(2, N= 185) = 26.6$ $p < 0.001$ [b,a]
Postpartum blues (no/yes)	151/93	8/20	7/26	$X^2(2, N= 305) = 27.7$ $p < 0.001$ [b,1,2]	102/45	0/16[c]	9/21	$X^2(2, N= 193) = 39.4$ $p < 0.001$ [b,a]
Stressful life events								
(number)	0.81 ± 1.27	1.46 ± 1.71	1.19 ± 1.18		1 ± 1.35	2.44 ± 2.06	1.97 ± 1.88	
(no/yes)	144/103	11/17	12/20	$X^2(2, N= 307) = 7.78$ $p = 0.020$	71/74	2/14	8/22	$X^2(2, N= 191) = 11.5$ $p = 0.003$ [b,1]
Breastfeeding T4 (no/yes)	63/182	14/14	8/25	$X^2(2, N= 306) = 7.62$ $p = 0.022$ [b,1]	41/96	9/7	9/20	$X^2(2, N= 182) = 4.56$ $p = 0.10$
Hamilton Depression Rating Scale T4	--	13.2 ± 2.88 N = 27	--	--	--	14.6 ± 4.18 N = 16	--	--

AD adjustment disorder, *HC* healthy controls, *PMS* premenstrual syndrome, *PPD* postpartum depression.

[a]No statistical analysis possible due to low expected cell counts.

[b]Bonferroni-corrected significant difference (*p* < 0.05) between [1] HC and PPD, [2] between HC and AD and/or [3] between PPD and AD.

[c]Significant group difference (*p* < 0.05) between first and validation cohort.

Univariate analyses of the first cohort

All data were analyzed using MATLAB R2018a, Python Jupyter Notebook 5.6.0, IBM SPSS Statistics 22 and jamovi 1.0.5.0[22]. Chi-square tests were performed to compare categorical sociodemographic-anamnestic variables across the groups in the first cohort. For continuous variables, logistic regressions were computed. Weekly mood and stress levels were calculated by averaging the corresponding bi-daily assessments. Mood-stress difference scores were calculated as the difference between both z-transformed variables to estimate individual discrepancies between perceived stress and mood (i.e., z-score mood minus z-score stress). Changes from baseline and the preceding week were computed for these variables. Dynamic changes in mood, stress, mood-stress difference, MPAS, and EPDS were analyzed using mixed effects repeated-measures analyses of variance (ANOVA) with week as within-subject and group as between-subject variable including an interaction term. Only post-hoc pairwise group comparisons (i.e., chi-square tests for categorical and binomial logistic regression for continuous sociodemographic-anamnestic variables, and independent samples t tests for mixed effects repeated-measures ANOVAs) were corrected for multiple testing using Bonferroni correction. The sample size was calculated as adequate for all univariate tests with a power of 0.8 and small to moderate effect sizes. Receiver operating characteristic (ROC) curves and their associated area under the curve (AUC) (within-sample) for differentiation between the three groups were computed for each measure per week.

Identification of most predictive combinations in the first cohort

Next, we aimed to evaluate if and which combinations of sociodemographic and clinical-anamnestic factors, mood, stress, MPAS and EPDS allow for an accurate differentiation between HC, PPD and AD in the first cohort. To that end, we used a logistic regression classifier (MATLAB built-in mnrfit and mnrval functions, no parameter optimization needed) performing 1000 repetitions of strict threefold cross-validation. The classification was performed for each pair-wise group comparison separately and oversampling was applied to the PPD and AD groups. Low-variance variables (family status, breastfeeding T0, education, completed professional education, income, and psychiatric diagnosis in previous pregnancy), i.e., variables with low group cell counts (less than 80% of expected cell counts >5), were excluded from the analysis in the whole sample (see Table 1 and Table S1). Independent samples t tests were performed in the training data to select the baseline variables to be included in the classifier ($p < 0.05$).

To identify the most sensitive combinations for early identification of PPD, the following nine feature combinations were evaluated: [1] baseline sociodemographic-anamnestic data alone, [2] mood scores, [3] stress scores, [4] mood-stress difference scores, [5] mood scores incl. changes (change to baseline and to preceding week), [6] stress scores incl. change scores, [7] mood-stress difference scores incl. changes, [8] combination of mood and stress scores incl. changes, [9] and combination of mood, stress, and mood-stress difference scores incl. changes. Combinations [1] to [9] were evaluated either alone or in combination with EPDS scores, MPAS scores or both. In addition, all combinations with features [2] to [9] were evaluated with and without inclusion of baseline sociodemographic-anamnestic information. The baseline sociodemographic-anamnestic information alone (i.e., feature combination [1]) served as null model for comparison with best performing models.

Balanced accuracies, sensitivities, specificities, positive and negative predictive values as well as ROC curves including the AUC were computed. The best performing combination (high balanced accuracy at earliest possible time-point) for each pair-wise comparison was selected for replication analysis. A logistic regression was computed for the selected combination using all participants. These results of the first cohort along with the validation plan were pre-registered on https://osf.io/ecmrp? view_only=6feb8e89818445a0b675621c8f22ba82. The obtained coefficients were applied to the prospectively collected validation cohort.

Application to the validation cohort

The selected and preregistered model as trained on the first dataset was then used to predict diagnoses in the independent validation cohort (Table S2). The class probability p for the validation cohort was obtained using the following standard logistic regression formula, where β denotes the coefficients and X the included features:

$$p = \frac{1}{1 + e^{-X\beta}}$$

As for the validation cohort, we computed balanced accuracy, sensitivity, specificity, AUC, ROC, and positive and negative predictive value by comparing predicted versus actual group labels. To obtain a chance level spread estimate for the classifier, we randomly permuted the "predicted" labels 1000 times across the validation cohort recomputing all performance measures and their 95% confidence interval.

Results

Sociodemographic-anamnestic and baseline group comparisons

In the first cohort, PPD and AD were associated with personal ($p < 0.001$ for HC vs. PPD and HC vs. AD) and familial psychiatric history ($p = 0.036$ for HC vs. PPD, $p = 0.009$ for HC vs. AD), subjective birth-related psychological traumas ($p = 0.024$ for HC vs. PPD, $p < 0.001$ for HC vs. AD), and postpartum blues ($p = 0.003$ for HC vs. PPD, $p < 0.001$ for HC vs. AD) (Table 1, S1 and S2). A higher PMS prevalence ($p = .012$ for HC vs. PPD) and reduced breastfeeding at T4 were observed in PPD compared to HC ($p = 0.021$). No differences were seen between PPD and AD. Similar effects were observed in the validation cohort for all sociodemographic-anamnestic factors (Table 1, Table S1; for odds ratios see Table S3).

Univariate analyses of the first cohort

The average participation over a total of 84 days of observation was 40 responses with a maximum of 45 responses, with no significant differences between the subsamples (HC: M = 40, max = 45; AD: M = 40, max = 44; PPD: M = 40, max = 45; F(2, 305) = 0.33, $p = 0.717$). Both PPD and AD showed a distinct pattern in weekly mood, stress, and mood-stress difference scores over the course of 12 weeks (significant time by diagnosis interactions – mood: F(13.8,1303) = 16.3, $p < 0.001$; stress: F(11.3,1026) = 9.85, $p < 0.001$; mood-stress difference: F(13.1,1162) = 17.3, $p < 0.001$) (Fig. 1A-C). The groups differed significantly in mood and mood-stress difference at all weeks ($p = 0.004$ for mood-stress baseline, all other $p < 0.001$) (see Tables S4 and S5). For stress, the difference was significant at all weeks except for baseline (all $p < 0.001$, see Table S6).

PPD had significantly lower mood levels compared to HC at all weeks except for baseline (Fig. 1A). AD had significantly lower mood relative to HC from baseline until week 6 reaching the highest difference at week 2. PPD had lower mood compared to AD from week 4 through week 12. Stress levels were significantly higher in PPD compared to HC from week 2 through week 12 and compared to AD between week 5 and week 12. AD had higher stress levels relative to HC from week 1 until week 4 (Fig. 1B). Mood-stress difference differed significantly between HC and PPD from week 1 through week 12, between HC and AD from week 1 through week 6, and between PPD and AD from week 4 through week 12 (Fig. 1C).

Both EPDS and MPAS showed significant time by diagnosis interactions (EPDS: F(6.87,1034) = 34.4, $p < 0.001$; MPAS: F(5.35,805) = 8.24, $p < 0.001$) with a significant between-group difference at all weeks (all $p < 0.001$) (Fig. 1D, E). EPDS scores were significantly lower in HC compared to PPD and AD at all time-points (T0-T4) ($p < 0.001$). The difference between PPD and AD was significant from T2 until T4 with higher EPDS scores in PPD women ($p < 0.001$). MPAS scores were significantly lower at all time points (T1–T4) in PPD ($p < 0.001$) and AD ($p < 0.001$ for T1-T3, $p = 0.008$ for T4) compared to HC. Lower MPAS scores were observed in PPD compared to AD at T4 ($p = 0.001$).

Prediction in the first cohort

Next, we evaluated which combinations of sociodemographic-anamnestic, mood, stress, EPDS, and MPAS data allow for reliable differentiation between PPD, AD, and HC. The outcomes of all evaluated combinations are summarized in Tables S7–14. For differentiation of PPD from HC, a high balanced accuracy of 87% was achieved at week 3 using a combination of baseline EPDS and follow-up EPDS and mood levels at week 3 (Table 2, Fig. 2A, and Table S7). The best early differentiation between AD and HC with a 91% balanced accuracy was also achieved at week 3 using a combination of baseline EPDS and follow-up EPDS, MPAS and mood scores at week 3 (Table 2, Fig. 2B, and Table S8). A reasonable differentiation of AD and PPD with a balanced accuracy of 76% was only achieved at week 6 using only the mood levels (Table 2, Fig. 2C, and Table S9). Logistic regression coefficients were trained with these combinations using the first cohort and applied to predict the diagnostic labels in the validation cohort (Table S2). The null model (i.e. sociodemographic-anamnestic information alone) performed inferior compared to the best performing models for all group comparisons (HC-PPD: BA = 0.72, HC-AD: BA = 0.75, AD-PPD: BA = 0.48; Table S9, Feature Combination 1).

Prediction in the validation cohort

The validation cohort had an average participation of 37 responses with a maximum of 45 responses for the remote assessments with no differences between the subgroups (HC: M = 38, max = 45; AD: M = 38, max = 43; PPD: M = 34, max =43; F(2, 190) = 1.51, $p = 0.223$). The classifier trained on the first cohort for differentiation of HC and PPD reached a high balanced accuracy of 93% in the validation cohort with a sensitivity of 88% and specificity of 99% (Table 2, Fig. 2D). The classifier differentiating HC and AD reached a balanced accuracy of 79% with a high specificity (98%) but only moderate sensitivity (60%) (Table 2, Fig. 2E). For PPD and AD differentiation, the selected classifier reached a balanced accuracy of 73%, again with high specificity (90%) but only low sensitivity (56%) (Table 2, Fig. 2F).

Discussion

Here, we adopted a within- and out-of-sample validation study design to identify combinations of
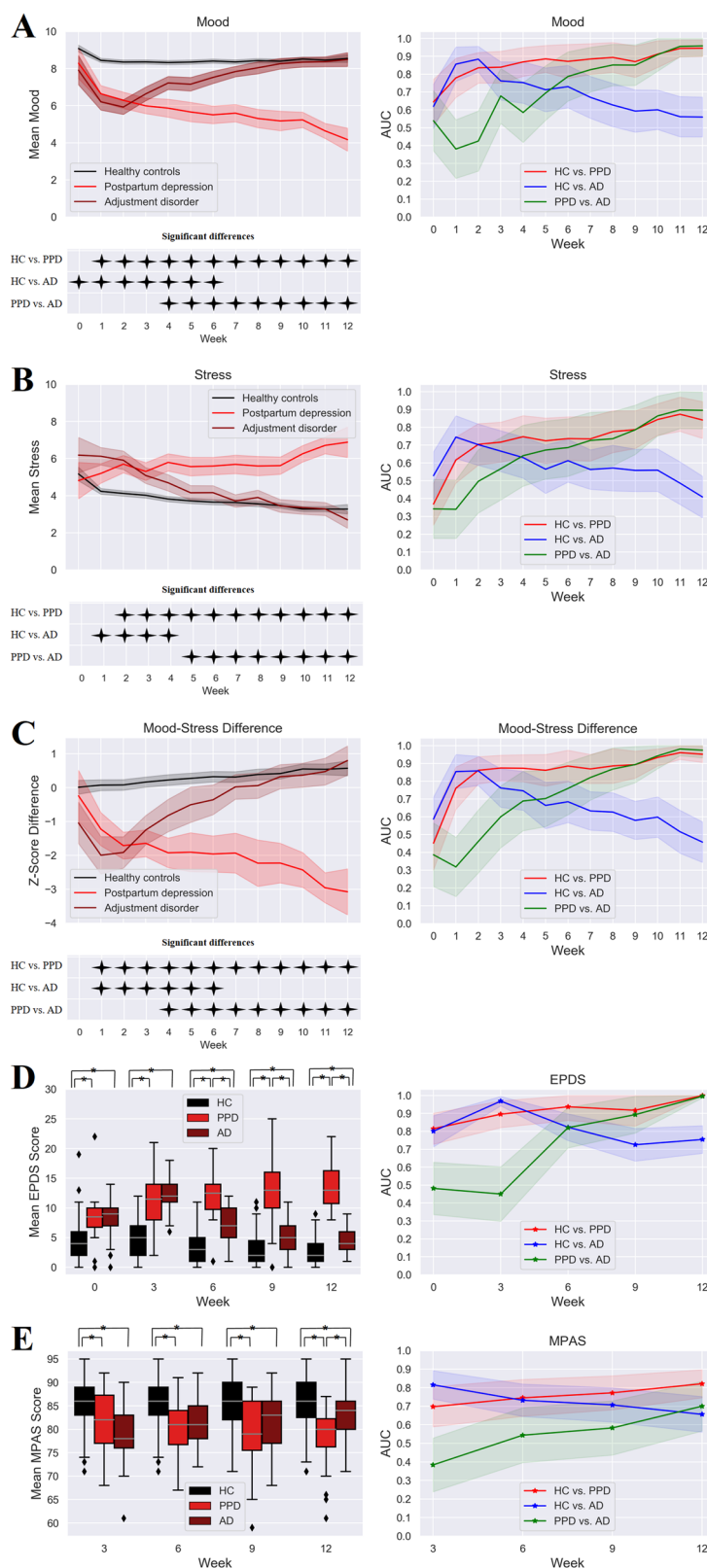
Fig. 1 Mood, stress, mood-stress difference, EPDS, and MPAS scores. Weekly mood (**A**), stress, (**B**) and mood-stress difference scores (**C**) incl. 95% confidence intervals, results of the simple effects analyses, and within-sample AUCs incl. 95% confidence interval for each group comparison. EPDS (**D**) and MPAS (**E**) mean scores and associated within-sample AUCs for each time point and group separately incl. their standard error and 95% confidence interval. Statistically significant *t* tests for group comparisons are marked with *.

Table 2 Results of prediction for the first and validation cohort.

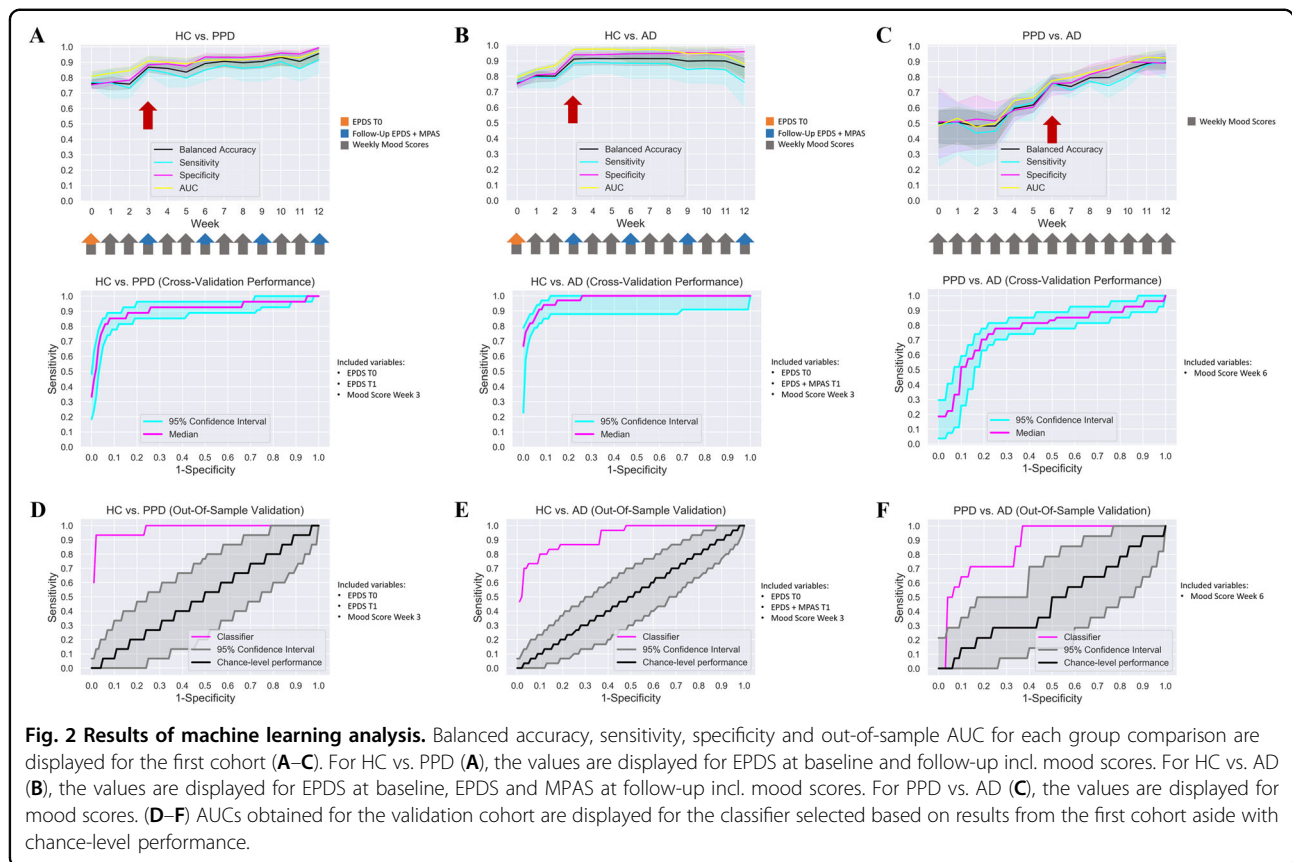|	HC vs. PPD			HC vs. AD			PPD vs. AD		
	Cross-validation (1st cohort)	Out-of-sample validation	Chance (95% CI)	Cross-validation (1st cohort)	Out-of-sample validation	Chance (95% CI)	Cross-validation (1st cohort)	Out-of-sample validation	Chance (95% CI)
Balanced accuracy	0.87	0.93	0.50 (0.44; 0.59)	0.91	0.79	0.50 (0.45; 0.57)	0.76	0.73	0.50 (0.35; 0.67)
Sensitivity	0.85	0.88	0.11 (0.00; 0.27)	0.88	0.60	0.12 (0.03; 0.23)	0.76	0.56	0.28 (0.07; 0.50)
Specificity	0.89	0.99	0.90 (0.89; 0.92)	0.94	0.98	0.88 (0.86; 0.90)	0.76	0.90	0.73 (0.63; 0.83)
PPV	0.45	0.88	0.10 (0.00; 0.25)	0.66	0.86	0.18 (0.05; 0.33)	0.66	0.75	0.32 (0.08; 0.58)
NPV	0.98	0.99	0.91 (0.89; 0.92)	0.98	0.92	0.83 (0.81; 0.85)	0.79	0.79	0.68 (0.59; 0.78)
AUC	0.91	0.98	0.50 (0.34; 0.65)	0.97	0.92	0.50 (0.39; 0.62)	0.79	0.88	0.50 (0.31; 0.68)

Balanced accuracy, sensitivity, specificity, positive and negative predictive value obtained for the validation cohort (out-of-sample validation). For comparison, results of the first cohort aside with chance-level classification for the validation cohort are also reported.
AD adjustment disorder, AUC area under the curve, CI confidence interval, HC healthy controls, PPD postpartum depression, PPV positive predictive value, NPV negative predictive value.

sociodemographic-anamnestic and clinical factors allowing for early and accurate identification and differentiation of PPD and AD in two large cohorts of postpartum women. In both cohorts high accuracy was achieved at week 3 for identification of PPD and AD compared to HC using a simple combination of EPDS, mood, and MPAS (for AD) assessments. In contrast, differentiation of PPD and AD was possible only from week 6 based solely on mood levels.

In both cohorts, the prevalence of PPD was slightly lower than the 10–20 % reported in the literature[23,24]. As the focus of our study was on prediction of PPD, we purposely excluded women with manifest depression at the time of inclusion in the study, which may explain the lower prevalence. Furthermore, studies estimating early prevalence of PPD may have included women with AD. Although there is an increased risk for PPD within the first postpartum year[25], meaning that some women may develop PPD after four to six weeks (i.e. late onset), this was not the case for our sample. In line with previous research, we found postpartum blues, psychiatric history, subjective birth-related psychological traumas, and PMS to be significant risk factors for PPD[14,26,27].

Interestingly, no differences between the PPD and AD groups were found with respect to risk factors, suggesting that similar mechanisms may be involved in the generation of initial depressive symptoms in both groups. Over the observation period, stress levels continuously increased in women with PPD whilst they normalized after about five weeks in AD. Descriptively, mood levels in AD followed the stress levels normalizing only after about seven weeks. The temporal delay is in line with the interpretation that reductions in stress may contribute to the recovery observed in mood. The increase in stress levels and the simultaneous decline in mood levels in PPD may indicate the contribution of stress-mediated components in line with previous studies reporting parenting stress among the most important postpartum factors[28,29]. Whilst not a causal factor on its own, parenting stress is likely to increase vulnerability to depression in high-risk individuals.

Similarly, PPD and AD displayed distinct temporal courses of EPDS and attachment scores as measured by MPAS. The EPDS temporal dynamics were highly similar to the observed stress and mood levels. The initially lowest attachment scores were found to increase in AD while PPD maintained the low attachment levels throughout the study. These observations underscore the necessity of longitudinal monitoring of both measures to better characterize the dynamic relationship between depressed mood and maternal attachment[30,31]. Differences in MPAS and EPDS remained significant between AD and HC at all time points. According to recent findings, child neurodevelopment is affected by maternal

Fig. 2 Results of machine learning analysis. Balanced accuracy, sensitivity, specificity and out-of-sample AUC for each group comparison are displayed for the first cohort (**A**–**C**). For HC vs. PPD (**A**), the values are displayed for EPDS at baseline and follow-up incl. mood scores. For HC vs. AD (**B**), the values are displayed for EPDS at baseline, EPDS and MPAS at follow-up incl. mood scores. For PPD vs. AD (**C**), the values are displayed for mood scores. (**D**–**F**) AUCs obtained for the validation cohort are displayed for the classifier selected based on results from the first cohort aside with chance-level performance.

depressive symptoms even when they do not exceed clinical thresholds[32,33]. Our observations emphasize the need for further detailed evaluation of potential consequences also for the AD group.

A combination of baseline EPDS and week 3 remote follow-up EPDS, and mood scores achieved about 90% balanced accuracy for early identification of PPD as compared to HC. The same combination with addition of MPAS achieved a similar accuracy for early identification of AD. Both findings were largely confirmed in the validation cohort with an accuracy reduction from 90 to 80% seen only for differentiation of AD and HC. None of the evaluated combinations allowed for an accurate early differentiation between PPD and AD with all classifiers performing close to chance level until week 5. A reasonable differentiation of both groups was only achieved through mood scores at week 6 with a moderately high accuracy but a high specificity for PPD as confirmed in the validation cohort. Our classification results suggest that a simple stepwise procedure including remote mood, EPDS, and MPAS assessments may be a promising approach towards early identification of PPD. Whilst week 3 remote testing provided a high accuracy and a particularly high specificity for detection of both populations at risk, week 6 data additionally allowed for further differentiation

between PPD and AD. In particular, the addition of mood scores led to a substantial increase in balanced accuracies for all group differentiations compared to all other feature combinations (e.g., addition of stress scores). Interestingly, the classifiers performed superior for the out-of-sample prediction in several cases. As we applied a strict cross-validation procedure the differences in prediction may simply reflect random variation in the accuracy of our model.

Three potential limitations need to be mentioned. First, as we did not register the reason for refusal during recruitment, we cannot exclude a bias based on the differences between women willing and women unwilling to participate. However, according to a recent study, there are no differences in motivation and willingness to participate between healthy controls and patients with psychiatric mood disorders[34]. Therefore, we do not expect any significant bias regarding the exclusion of women with PPD or AD based on their refusal to participate in the study. A potential bias introduced by the recruitment after childbirth vs. before childbirth may be a second limitation. However, the main goal of the current study was the identification of a risk group through a method, which could be easily applied in routine care. Prediction before childbirth may be more difficult to incorporate into

routine care as it may require the transfer of information between multiple institutions (e.g. gynecologist and hospital). Third, oversampling was applied only to the cross-validation in the first cohort, but not to the training of the classifier for prediction in the validation cohort, resulting in a potential bias of the logistic regression classifier due to asymmetric group sizes. However, considering that the highly similar results for the cross-validation and the out-of-sample (with the out-of-sample validation results being even superior at times), these findings indicate a minor influence of the asymmetric group sizes on the outcomes of our study.

In summary, by means of a longitudinal approach we identify and validate combinations of remote assessments allowing for early and accurate identification and differentiation of PPD and AD using a step-wise procedure. By administering the EPDS and mood assessments in-clinic immediately after childbirth and a second assessment remotely after three weeks, these findings can be easily translated into routine care. The behavioral and clinical time courses over 12 weeks provided important insight into the development and interaction of mood, stress, and maternal sensitivity in the first weeks postpartum.

Author details
[1]Institute of Neuroscience and Medicine, Brain & Behaviour (INM-7), Research Centre Jülich, Jülich, Germany. [2]Institute of Systems Neuroscience, Medical Faculty, Heinrich Heine University Düsseldorf, Düsseldorf, Germany. [3]Department of Psychiatry, Psychotherapy and Psychosomatics, Medical Faculty, Uniklinik RWTH Aachen University, Aachen, Germany. [4]Institute of Neuroscience and Medicine, Brain Structure Function Relationship (INM-10), Research Centre Jülich, Jülich, Germany. [5]Department of Gynecology and Obstetrics, Medical Faculty, Uniklinik RWTH Aachen University, Aachen, Germany. [6]Department of Obstetrics, RoMed Hospital Rosenheim, Rosenheim, Germany. [7]Clinical Institute of Psychosomatic Medicine and Psychotherapy, Medical Faculty, Heinrich-Heine-Universität Düsseldorf, Düsseldorf, Germany

Author contributions
L.H. performed all analyses and wrote the manuscript. S.B.E., N.C., and J.D. designed the overall study. N.C., P.S., U.H., E.S., T.W.G., and S.S. conducted the clinical studies. All authors reviewed and commented on the manuscript.

Data availability
The data of this study are not publicly available due to privacy and ethical restrictions. Data to support the findings of this study are available upon reasonable request.

Code availability
The computer code used for the prediction analysis is available upon reasonable request.

Conflict of interest
J.D. is a former employee and current consultant for F.Hoffmann-La Roche. All authors report no conflicts of interest with respect to the work presented in this study.

References
1. Munk-Olsen, T., Laursen, T. M., Pedersen, C. B., Mors, O. & Mortensen, P. B. New parents and mental disorders: a population-based register study. *JAMA* **296**, 2582–2589 (2006).
2. Stewart, D. E. et al. Postpartum depression: Literature review of risk factors and interventions. *Tor. Univ. Health Netw. Women's Health Program Tor. Public Health* 1–289 (2003).
3. American PsychiatricAssociation. *Diagnostic and statistical manual of mental disorders (DSM-5®)*. (American Psychiatric Pub, 2013).
4. Halfin, A. Depression: the benefits of early and appropriate treatment. *Am. J. Manag. Care* **13**, S92–S97 (2007).
5. Shen, H. et al. Associations of parental depression with child school performance at age 16 years in Sweden. *JAMA Psychiatry* **73**, 239–246 (2016).
6. Weissman, M. M. et al. Offspring of depressed parents: 20 years later. *Am. J. Psychiatry* **163**, 1001–1008 (2006).
7. Pearson, R. M. et al. Maternal perinatal mental health and offspring academic achievement at age 16: the mediating role of childhood executive function. *J. Child Psychol. Psychiatry* **57**, 491–501 (2016).
8. Yang, J. et al. Effects of parental emotional warmth on the relationship between regional gray matter volume and depression-related personality traits. *Soc. Neurosci.* **12**, 337–348 (2017).
9. Weissman, M. M. et al. Treatment of maternal depression in a medication clinical trial and its effect on children. *Am. J. Psychiatry* **172**, 450–459 (2015).
10. O'Hara, M. W. Postpartum depression: what we know. *J. Clin. Psychol.* **65**, 1258–1269 (2009).
11. O'Hara, M. W. & McCabe, J. E. Postpartum depression: current status and future directions. *Annu. Rev. Clin. Psychol.* **9**, 379–407 (2013).
12. Galea, L. A. & Frokjaer, V. G. Perinatal depression: embracing variability toward better treatment and outcomes. *Neuron* **102**, 13–16 (2019).
13. Schipper-Kochems, S. et al. Postpartum depressive disorder–psychosomatic aspects. *Geburtshilfe Frauenheilkd.* **79**, 375–381 (2019).
14. Buttner, M. M. et al. Examination of premenstrual symptoms as a risk factor for depression in postpartum women. *Arch. Women's Ment. Health* **16**, 219–225 (2013).
15. Tortajada, S. et al. Prediction of postpartum depression using multilayer perceptrons and pruning. *Methods Inf. Med.* **48**, 291–298 (2009).
16. Jiménez-Serrano, S., Tortajada, S. & García-Gómez, J. M. A mobile health application to predict postpartum depression based on machine learning. *Telemed. E-Health* **21**, 567–574 (2015).
17. Hamilton, M. A rating scale for depression. *J. Neurol. Neurosurg. Psychiatry* **23**, 56 (1960).
18. Steiner, M., Haskett, R. F. & Carroll, B. J. Premenstrual tension syndrome: the development of research diagnostic criteria and new rating scales. *Acta Psychiatr. Scand.* **62**, 177–190 (1980).
19. Goodman, L. A., Corcoran, C., Turner, K., Yuan, N. & Green, B. L. Assessing traumatic event exposure: general issues and preliminary findings for the stressful life events screening questionnaire. *J. Trauma. Stress. Publ. Int. Soc. Trauma. Stress Stud.* **11**, 521–542 (1998).
20. Bergant, A. M., Nguyen, T., Heim, K., Ulmer, H. & Dapunt, O. German language version and validation of the Edinburgh postnatal depression scale. *Dtsch. Med. Wochenschr. 1946* **123**, 35–40 (1998).
21. Condon, J. T. & Corkindale, C. J. The assessment of parent-to-infant attachment: Development of a self-report questionnaire instrument. *J. Reprod. Infant Psychol.* **16**, 57–76 (1998).
22. The jamovi project (2020). jamovi (Version 1.2) [Computer Software]. Retrieved from https://www.jamovi.org.
23. Anokye, R., Acheampong, E., Budu-Ainooson, A., Obeng, E. I. & Akwasi, A. G. Prevalence of postpartum depression and interventions utilized for its management. *Ann. Gen. Psychiatry* **17**, 18 (2018).

24. Hahn-Holbrook, J., Cornwell-Hinrichs, T. & Anaya, I. Economic and health predictors of national postpartum depression prevalence: a systematic review, meta-analysis, and meta-regression of 291 studies from 56 countries. *Front. Psychiatry* **8**, 248 (2018).

25. Stowe, Z. N., Hostetter, A. L. & Newport, D. J. The onset of postpartum depression: Implications for clinical screening in obstetrical and primary care. *Am. J. Obstet. Gynecol.* **192**, 522–526 (2005).

26. Beck, C. T. Predictors of postpartum depression: an update. *Nurs. Res.* **50**, 275–285 (2001).

27. O'Hara, M. W. & Swain, A. M. Rates and risk of postpartum depression—a meta-analysis. *Int. Rev. Psychiatry* **8**, 37–54 (1996).

28. Venkatesh, K. K., Phipps, M. G., Triche, E. W. & Zlotnick, C. The relationship between parental stress and postpartum depression among adolescent mothers enrolled in a randomized controlled prevention trial. *Matern. Child Health J.* **18**, 1532–1539 (2014).

29. Leigh, B. & Milgrom, J. Risk factors for antenatal depression, postnatal depression and parenting stress. *BMC Psychiatry* **8**, 24 (2008).

30. Akman, İ. et al. Breastfeeding duration and postpartum psychological adjustment: role of maternal attachment styles. *J. Paediatr. Child Health* **44**, 369–373 (2008).

31. Mason, Z. S., Briggs, R. D. & Silver, E. J. Maternal attachment feelings mediate between maternal reports of depression, infant social–emotional development, and parenting stress. *J. Reprod. Infant Psychol.* **29**, 382–394 (2011).

32. Goodman, S. H. et al. Maternal depression and child psychopathology: A meta-analytic review. *Clin. Child Fam. Psychol. Rev.* **14**, 1–27 (2011).

33. Evans, J. et al. The timing of maternal depressive symptoms and child cognitive development: a longitudinal study. *J. Child Psychol. Psychiatry* **53**, 632–640 (2012).

34. Morán-Sánchez, I., Maurandi-López, A. & Pérez-Cárceles, M. D. Assessment of motivations and willingness to participate in research of outpatients with anxiety, mood, and psychotic disorders. *J. Empir. Res. Hum. Res. Ethics* **13**, 546–560 (2018).