



HHS Public Access

Author manuscript

Int J Eat Disord. Author manuscript; available in PMC 2016 September 01.

Published in final edited form as:

Int J Eat Disord. 2015 September ; 48(6): 654–662. doi:10.1002/eat.22361.

Psychosocial Factors Associated with Bulimia Nervosa during Pregnancy: An Internal Validation Study

Hunna J. Watson, PhD^{1,2,3,4,5,*}, Ann Von Holle, MS¹, Cecilie Knoph, MA⁶, Robert M. Hamer, PhD^{1,7}, Leila Torgersen, PhD⁶, Ted Reichborn-Kjennerud, MD^{6,8}, Camilla Stoltenberg, MD^{9,10}, Per Magnus, MD, PhD⁹, and Cynthia M. Bulik, PhD^{1,11,12}

¹Department of Psychiatry, University of North Carolina at Chapel Hill, Chapel Hill, USA

²Eating Disorders Program, Child and Adolescent Health Service, Department of Health in Western Australia, Perth, Australia

³Centre for Clinical Interventions, Department of Health in Western Australia, Perth, Australia

⁴School of Paediatrics and Child Health, The University of Western Australia, Perth, Australia

⁵School of Psychology and Speech Pathology, Curtin University, Perth, Australia

⁶Division of Mental Health Services, Akershus University Hospital, Lørenskog, Norway

⁷Department of Biostatistics, University of North Carolina at Chapel Hill, Chapel Hill, USA

⁸Institute of Clinical Medicine, University of Oslo, Oslo, Norway

⁹Division of Epidemiology, Norwegian Institute of Public Health, Oslo, Norway

¹⁰Department of Public Health and Primary Health Care, University of Bergen, Oslo, Norway

¹¹Department of Nutrition, University of North Carolina at Chapel Hill, Chapel Hill, USA

¹²Department of Medical Epidemiology and Biostatistics, Karolinska Institutet, Stockholm, Sweden

Abstract

Objective—The aim of this paper was to internally validate previously reported relations (1) between psychosocial factors and bulimia nervosa (BN) outcomes during pregnancy.

Method—This study is based on the Norwegian Mother and Child Cohort Study (MoBa) conducted by the Norwegian Institute of Public Health. Participants were women enrolled during pregnancy ($N = 69,030$). Internal validity was evaluated by way of bootstrapped parameter estimates using the overall sample and a split sample calibration approach.

Results—Bootstrap bias estimates were below the problematic threshold, and extend earlier findings(1) by providing support for the validity of the models at the population level of all

*Correspondence to: Dr Hunna Watson, Department of Psychiatry, University of North Carolina at Chapel Hill, CB#7160, Chapel Hill, NC, 27599-7160, USA. hunna_watson@med.unc.edu.

Declaration of Interest

Dr. Bulik is a Consultant for Shire Pharmaceuticals. Other authors report no conflicts of interest. The authors alone are responsible for the content and writing of the paper.

pregnant women in Norway. Bootstrap risk ratios indicated that prevalence, incidence, and remission of BN during pregnancy were significantly associated with psychosocial factors. The split sample procedure showed that the models developed on the training sample did not predict risks in the validation sample.

Discussion—This study characterizes associations between psychosocial exposures and BN outcomes among pregnant women in Norway. Women with lifetime and current self-reported psychosocial adversities were at a much higher risk for BN during pregnancy. Psychosocial factors were associated with BN remission during pregnancy, inviting the prospect of enhancing therapeutic interventions. We consider the findings in the context of reproducibility in science.

Keywords

bulimia nervosa; course; eating disorders; incidence; internal validation; MoBa; pregnancy; The Norwegian Mother and Child Cohort Study

Introduction

Pregnancy is a time of social, psychological, and physical change, and can be a turning point for recovery from eating disorders (EDs) or adversely, for onset or relapse.(2, 3) EDs prior to and during pregnancy increase the risk of pregnancy complications and negative birth outcomes, such as miscarriages, fetal growth problems, perinatal mortality, low or high birth weight, premature birth, and birth defects, although adequate weight gain in anorexia nervosa (AN) seems to buffer against adverse outcomes.(4-6) Mothers with bulimia nervosa (BN) self-report different feeding styles (i.e., restrictive) and greater infant feeding problems compared with mothers without EDs,(7) which may influence infant development. Identification of EDs can be incorporated into routine obstetric assessment, and pregnancy may be an opportune window for engaging women in ED treatment given motivation to enhance baby health.

The Norwegian Mother and Child Cohort Study (MoBa) is a prospective population-based pregnancy cohort study, which recruited over 100,000 pregnancies between 1999 and 2009, and includes information on EDs in the six months prior to and during pregnancy. Pre-pregnancy, approximately 1 in 100 women in MoBa met criteria for BN, and during pregnancy the most common outcomes were partial remission (33%) and remission (37%). (8) BN onset during pregnancy was rare (0.1%). The mechanisms that influence the course of illness are poorly understood and a thorough understanding carries potential for assertive and targeted interventions.

An initial planned, previously published(1) analysis of MoBa data by our group approximately halfway into cohort recruitment ($N = 41,157$) investigated psychosocial exposures related to BN outcomes during pregnancy. Remission was associated with higher self-esteem, higher life satisfaction, and lower anxiety and depression during pregnancy. Incidence during pregnancy was associated with higher anxiety and depression, lower self-esteem and lower life satisfaction, and continuation and partial remission had no associations with psychosocial exposures.(1) Prevalence of BN during pregnancy was predicted by many psychosocial exposures. To date, this is the only study that has quantified

the relations between psychosocial exposures and BN during pregnancy and many significant associations were observed.

Few opportunities exist for external replication of Knoph Berg and colleagues' findings in independent samples. The challenge impeding scientific progress in research on predictors of course of eating disorders in pregnancy is the requirement for a sufficiently large population-based sample. This is needed to give numerically respectable sample size splits by diagnosis and each possible outcome (i.e., incidence, continuation, etc.). Internal validation, sometimes termed quasi-replication,(9) is another approach to assessing the reproducibility of findings and evaluates whether findings in a sample generalize to independent cases in the same cohort (i.e., split-half validation) or to the wider true population (i.e., bootstrap estimates of bias).(10)

At the time of the initial analysis, MoBa was approximately halfway toward their recruitment goals. Our a priori statistical analysis plan called for the internal validation of the models. Hence, the aim of this study was to conduct an internal validation of the models established in the initial study. We hypothesized that the models would be internally valid.

Method

Design and participants

This study is nested within the MoBa study, a prospective population-based pregnancy cohort study conducted by the Norwegian Institute of Public Health.(11) The total sample comprised 69,030 pregnant women who were participants in the MoBa version 7 of the quality-assured data files (released for research on January 2013) and met inclusion requirements.

MoBa participants were recruited from all over Norway from 1999-2009 and 40.6% of invited women consented to participate (<http://www.fhi.no/moba-en>). The cohort now includes 114,500 children, 95,200 mothers and 75,200 fathers. MoBa is linked to Norwegian health registries, including the Medical Birth Registry of Norway (MBRN), and since MBRN was established in 1967, all stillbirths and live births in Norway after gestational week 12 require mandatory notification by midwives and doctors. This has probably resulted in minimal missed pregnancies.

Figure 1 contains a participant flow diagram of sample selection. The total sample size of the present study (69,030) is less than the overall number of mothers in the MoBa cohort (95,200) because inclusion and exclusion criteria were applied. The inclusion criteria were a) a first pregnancy during the study period [excluded 15,626 other pregnancies] b) had a singleton birth [excluded 4,130 multiples] and c) had a live birth [excluded 480 stillbirths], and resulted in 83,953 total included participants (some individuals met more than one criterion). The exclusion criteria were a) completed an early pilot version of the Questionnaire (Q1) survey [2,483] b) had BN pre-pregnancy but migrated to another ED diagnosis during pregnancy (not available in the context of this sample) c) weighed < 30 kg (67 lbs) and > 300 kg (661 lbs) before and during pregnancy [67] d) were < 100 cm (3.3 feet) [183] e) returned the survey after the birth of the child [260] f) had missing pregnancy

ID information [0] and g) had a missing age value [0]: there were 69,030 participants in total after these exclusion criteria were applied.

The overall sample ($N = 69,030$) completed Q1 at a median of 17.1 weeks gestation (interquartile range = 15.9 to 18.7 weeks). The average maternal age was 30.0 ($SD = 4.6$) years, 96.3% ($n = 66,280$) were married or cohabiting, and 56.4% ($n = 38,905$) were nulliparous.

Informed consent was obtained from each MoBa participant upon recruitment. The study was approved by The Regional Committee for Medical Research Ethics in South-Eastern Norway and the biomedical Institutional Review Board at the University of North Carolina at Chapel Hill.

Measures

Data for this study were derived from Moba Q1

(<http://www.fhi.no/dokumenter/1f32a49514.pdf>). Information on demographics was obtained from Q1 and the MBRN.

Eating disorders—Q1 contained items that assessed ED symptoms based on Diagnostic and Statistical Manual criteria.⁽¹²⁾ These criteria have been used in previous MoBa studies on EDs^(1, 13) and in the Norwegian Institute of Public Health Twin Panel.⁽¹⁴⁾

Q1 items were used to establish diagnostic algorithms as per the original investigation.⁽¹⁾ Broadly defined BN was operationalized as binge eating and either purging (vomiting, laxatives) or non-purging (exercise, fasting) compensatory behaviors occurring at least once per week. The binge eating item captured the two key features of binge eating: eating episodes that were unusually large and accompanied by a sense of loss of control. Purging was differentially assessed with regard to pregnancy-induced nausea and presented as a choice in an item about methods used to deliberately control weight. A diagnosis of broadly defined BN was not assigned if the individual met criteria for broadly defined AN also. Those described as having no eating disorder did not have a diagnosis of AN, BN, EDNOS-P, and BED, as operationalized previously.⁽¹⁾ Q1 assessed ED status in the six months prior to pregnancy (i.e., retrospectively) and at questionnaire administration (approximately 17.1 weeks gestation).

Continuation, remission, and partial remission of BN were only applicable to those who reported BN in the six months prior to pregnancy. For continuation, BN criteria were met during pregnancy. Remission was the absence of binge eating and purging and non-purging compensatory behaviors during pregnancy. Partial remission was the continued presence of binge eating but absence of purging and non-purging compensatory behaviors during pregnancy. Incidence was applicable only to those who reported no BN in the six months prior to pregnancy, and was assigned if the criteria for BN were met during pregnancy. Prevalence of BN was defined as reporting BN during pregnancy (continuation and incident cases). Lifetime BN was not assessed.

Psychological factors—A 5-item version of the Hopkins Symptom Checklist-25 assessed anxiety and depression symptoms in the previous two weeks on a Likert scale from 1 (not bothered) to 4 (very bothered).⁽¹⁵⁾ The short form correlates ($r = 0.92$) with the total score and has good psychometric properties.^(15, 16) Life satisfaction was assessed with the 5-item Satisfaction With Life Scale^(17, 18) which is valid, reliable, and has been used in hundreds of studies.⁽¹⁸⁾ Each item ranges from 1 (strongly disagree) to 7 (strongly agree). Self-esteem was assessed with a 4-item version of the Rosenberg Self-Esteem Scale (RSES).⁽¹⁹⁾ It correlates well ($r = 0.95$) with the original scale⁽²⁰⁾ and scores for each item range from 1 (strongly disagree) to 4 (strongly agree). Satisfaction with partner relationship was measured with the 10-item Relationship Satisfaction Scale,^(21, 22) developed within MoBa and partially based on the Marital Satisfaction Scale.⁽²³⁾ The scale has good psychometric properties and correlates 0.92 with the Quality of Marriage Index^(22, 24). Scale scores were derived by averaging the items for each respective scale.

Lifetime major depression was assessed using six questions designed to capture DSM-IV major depressive disorder including duration (i.e., two weeks). The self-report method has been used in previous research.⁽²⁵⁾

Adverse life events—Physical abuse was assessed during pregnancy, six months before pregnancy, and earlier in life (“Have you ever in your adult life been slapped, hit, kicked or bothered in any way physically?”), as was sexual abuse (“Have you ever been pressured or forced to have sexual intercourse?”). Time points were collapsed into single yes/no binary variable reflecting lifetime physical abuse or sexual abuse.

Health behaviors—Lifetime smoking was a binary (yes/no) variable. Frequency of alcohol consumption was measured with a dichotomous variable of drinking alcohol two times per week (“yes”) in the three months prior to pregnancy.

Statistical analysis

Bootstrapping and split-sample analysis⁽²⁶⁾ were used in the present study. These were the same methods used in the first MoBa validation study by our group.⁽²⁷⁾ The bootstrapping and split-sample methods assess internal validity but apply to different situations, so results from one approach are not directly comparable with the other.

Bootstrap estimates of bias characterize the extent to which the models are free of bias. The bootstrap risk ratio (RR) is the mean RR computed from resamples from the entire MoBa sample (69,030) and represents the estimate that would be obtained by taking many samples from the true population (i.e., pregnant women in Norway). The bias estimate equals the bootstrap RR minus the RR from the observed sample — the entire MoBa sample.^(28, 29) Resample RR estimates were generated using 1,000 samples obtained by repeated random sampling with replacement. These estimates were averaged to obtain the bootstrap RR and standard deviation, which was used to construct the 95% CI. Bias estimates of less than 0.25 standard errors are considered substantially small.⁽²⁸⁾ The false discovery rate (FDR) controlling procedure controlled for multiple testing.⁽³⁰⁾ Inferences made in the context of the bootstrap resampling approach rest on key assumptions (i.e., independence of samples).

The split-sample method evaluated the predictive ability of the models from the original study.(1) This method determines whether models developed on one sample fit when applied to a sample from the same cohort with no overlap of subjects. Participants from the entire MoBa cohort were split into a “training” sample ($n = 36,701$) based on participants in original study(1) and a remaining “validation” sample ($n = 32,329$) of participants not included in the original study (Figure 1). Incidentally, the training sample approximates but is not identical to the Knoph Berg et al. sample because MoBa datasets have unique pregnancy ID codes that preclude linking subjects across versions. Second, another factor is different in this analysis compared to the original one in Knoph Berg et al.:(1) we did not use the multiple imputation method to estimate model parameters so as to simplify the analyses. Both of these factors contribute to the estimates being different across what we call the “training” sample in this analysis and the original sample in Knoph Berg et al.(1) derived from the version 2 MoBa data release.

The split sample method uses a calibration procedure that estimates the degree of agreement between predicted frequencies derived from the model developed with the parameter estimates and frequencies observed in the validation sample (10, 31). The models from the training dataset provide the β parameters used in the calibration modeling.(10, 31) First, RR estimates of the five outcomes (i.e., incidence, remission, partial remission, continuation, prevalence) were estimated from Poisson regression models from the training dataset. Each model had a psychosocial variable for the independent variable, and was adjusted for maternal age and education as potential confounding factors. Each independent variable was modelled separately as per the original study.(1) The calibration procedure produces regression coefficients, which are used to test the null hypothesis of departure from perfect fit (a calibration slope of 1). In a perfectly calibrated model one would expect, for example, the predicted remission count to be the same as the observed remission count. A $\beta < 1$ provides evidence that the predicted risks using the training set model regression coefficients overestimate the observed risks in the validation data. The FDR procedure addressed multiple testing. Brier scores indicate how far the predicted estimates are from the observed estimates in the validation sample, estimates closer to 0 indicate better calibrated predictions.(32) Any model fitted in one sample is likely to fit less well in an independent sample, since the parameters in the model were selected to maximize a likelihood function in the initial sample. The original model is subject to random error sources, so the calibration process capitalizes on chance. The important question is whether the poorer fit is so poor that one can conclude that the initial fit in the training data capitalized on chance and has poor predictive ability when applied to an independent dataset.

Results

Descriptive statistics in Table 1 show the psychosocial characteristics for affected and unaffected women in the total sample. The unaffected women appeared to be more advantaged, with lesser experience of psychosocial risks and adverse life events.

Bootstrap method

The bootstrap estimates of the RR and 95% CI of all models are shown in Table 2, with the bias estimates. In all of the 44 models, bias did not exceed 0.25 of the standard error, suggesting excellent estimation. The bootstrap RRs indicate that several outcomes were significantly related to psychosocial exposures. Remission was positively associated with relationship satisfaction, self-esteem, and life satisfaction. Incidence was positively associated with physical abuse and anxiety and depression symptoms, and inversely associated with life satisfaction and self-esteem. Continuation and partial remission were not significantly related to any psychosocial factors and prevalence was associated with all psychosocial factors. In sum, the bootstrap analyses suggested that all models tested were low in bias and performed well, that is, the sample results (based on the entire MoBa sample) can be assumed to predict outcomes in the population well (i.e., all pregnant women in Norway).

Split-sample method

Figure 2 shows the results of the Poisson regression analyses with the RR estimates for the predictive models derived from the training sample split half. The results for the validation sample are plotted, too. There were several significant exposure-outcome associations (i.e., anxiety and depression symptoms with incidence).

Calibration models are shown in Table 3. We hypothesized that the predictive models fitted on the training sample would have good calibration properties, predicting outcomes in the validation sample. The performance of the training sample models, as evaluated with the overall β estimate, yielded mixed results regarding calibration and internal validity. Estimated probabilities of continuation across the independent variables showed no values significantly different from one. Thus, observed values in the validation data set are not significantly different from predicted values. However, some calibration values were quite different from one, despite a lack of statistical significance. For example, using life satisfaction as a predictor of continuation yielded a β of 0.46. While not significant, this value does not indicate good calibration. Brier scores also suggest departures from good calibration (see Table 4).

The outcome of incidence yielded calibration estimates significantly different from one, indicating poor agreement between the validation and training sample; one example is symptoms of anxiety and depression. The overall β for this calibration model is 0.33 indicating the log of the predicted RR is over 3.0 times greater than the observed RR. Unlike the prior example for continuation, this value is significantly different from one most likely because the sample is much larger for estimates of incidence compared to continuation. The Brier scores suggest reasonable performance of these models. The case for model performance when considering estimates of remission and partial remission is not much different. The calibration estimates are quite different from one, statistically significant for over half the independent variables, and Brier scores indicate poor performance. Prevalence as an outcome has a slightly better profile when comparing the overall β estimates to those for remission. Only three overall estimates are significantly different from one and many show overall estimates close to one.

When applying a conservative criterion to the β estimates in Table 3 of non-significant estimates that show less than a 10% ratio between measures of RRs, there are few examples. Using the same criterion for continuation, incidence, remission, and partial remission, only anxiety and depression symptoms and relationship satisfaction as predictors of continuation demonstrate evidence of good model calibration between predicted and observed probabilities.

In sum, the calibration statistics indicate that the probability models based on the training sample in general do not predict outcomes sufficiently well in the validation sample. In most cases, observed associations in the validation sample were overestimated, leading to higher predicted risks than observed.

Discussion

The aim of this study was to validate previously established models of the relation between psychosocial factors and BN during pregnancy in a large Norwegian population-based pregnancy cohort, MoBa. Our approach extends the work of Knoph Berg et al.(1) by using internal validation techniques to assess the performance of the models established in the original study. A first key finding of the present study is that the MoBa cohort validly reflects the degree to which psychosocial factors predict onset, remission, partial remission, continuation, and prevalence of BN among pregnant women in Norway. The second key finding is that in parallel with the original study,(1) the bootstrap RRs confirm several significant associations between psychosocial factors and course of illness of BN in pregnancy. These findings suggest that psychosocial factors may play a protective or risk role for BN among pregnant women.

Building on the original study, (1) the bootstrap analyses provide evidence that the models of associations between psychosocial factors and BN outcomes in the MoBa cohort have low bias and thus good internal validity. The findings suggest that over a decade-long period, pregnant women in Norway had a significantly higher risk of BN onset if they had higher anxiety and depression, lower life satisfaction, lower self-esteem, and a history of physical abuse. Women with BN pre-pregnancy were more likely to experience remission if they had higher relationship satisfaction, life satisfaction, and self-esteem. Partial remission and continuation of BN in pregnancy were not associated with any psychosocial factors, while prevalence of BN in pregnancy was associated with greater adversity on a host of psychosocial factors; anxiety and depression, life satisfaction, self-esteem, relationship satisfaction, lifetime major depression, sexual abuse, physical abuse, lifetime smoking, and alcohol use.

There was sufficiently convincing evidence from the bootstrap analyses and regression models in the validation sample (Figure 2) to suggest replication of the original study findings that psychosocial factors relate to BN outcomes in pregnancy. The failure to replicate within the split sample approach to internal validation implicates several factors; first, data splitting approaches are susceptible to differences due to chance when applying the split.(10) Second, unmeasured confounding differences between the two samples may contribute to instability in associations. This second explanation is unlikely for the present

study as other MoBa studies have not observed diversity(33, 34) and over time, recruitment has extended to the whole of Norway (50/52 hospitals). Third, the degree of random error, due to a smaller sample size when developing the models, is likely to have contributed to the lack of calibration. The split sample findings suggest that the magnitude of the risk associations in the original study did not carry over in the same manner to the independent sample from the same cohort. The quantified RRs derived from the training sample somewhat overestimated the likelihood of the BN outcomes based on the psychosocial factors. This implies that we cannot be confident of the RR value when projecting risk to individuals in a clinical context.

This study has limitations. Only correlational and not cause and effect relations between exposures and outcomes can be inferred due to the observational nature of the study. Diagnostic measures were self-report—a practical necessity given the large sample size—yet clinician-report is optimal. The pre-pregnancy items required recall up to nine months prior to assessment, the little data that exist suggest that eating disorder symptoms recalled up to 30 months after baseline correlate with baseline measurements(35), but this has only been examined for clinician-administered assessments. The questionnaire completed during pregnancy was filled out over a wide time range (4.0 to 41.6 weeks) with a relatively narrow interquartile range (15.9 to 18.7 weeks). Self-report data on sensitive psychosocial factors, such as smoking(36) and sexual abuse, may introduce underreporting bias. The MoBa participation rate is around 40%. Underrepresented individuals in MoBa include the youngest women (< 25 years), those living alone, and mothers with greater than two previous births or previous stillbirth. The large sample size, progression to a whole-of-country sampling frame, and generally synonymous background characteristics between the MoBa sample and all women giving birth in Norway suggest low selection bias and high national representativeness.

In conclusion, this study confirms the role of psychosocial factors in BN and course of illness during pregnancy. Specific knowledge about potentially modifiable risk and protective factors, and maintaining and prognostic factors, is a necessary step toward prevention and treatment of eating disorders in pregnant women.

Acknowledgements

We are grateful to all the participating families in Norway who take part in this ongoing cohort study.

Funding

The Norwegian Mother and Child Cohort Study is supported by the Norwegian Ministry of Health and the Ministry of Education and Research, NIH/NIEHS (contract no NO-ES-75558), NIH/NINDS (grant no.1 UO1 NS 047537-01 and grant no. 2 UO1 NS 047537-06A1), and the Norwegian Research Council/FUGE (grant no. 151918/S10). This research was supported by NIH grant R01 HD047186 (PI:Bulik).

References

1. Knoph Berg C, Bulik CM, Von Holle A, Torgersen L, Hamer R, Sullivan P, Reichborn-Kjennerud T. Psychosocial factors associated with broadly defined bulimia nervosa during early pregnancy: Findings from the Norwegian Mother and Child Cohort Study. *Australian and New Zealand Journal of Psychiatry*. 2008; 42:396–404. [PubMed: 18473258]

2. Blais MA, Becker AE, Burwell RA, Flores AT, Nussbaum KM, Greenwood DN, et al. Pregnancy: outcome and impact on symptomatology in a cohort of eating-disordered women. *International Journal of Eating Disorders*. 2000; 27:140–9. [PubMed: 10657887]
3. Tiller J, Treasure J. Eating disorders precipitated by pregnancy. *European Eating Disorders Review*. 1998; 6:178–87.
4. Micali N, Simonoff E, Treasure J. Risk of major adverse perinatal outcomes in women with eating disorders. *British Journal of Psychiatry*. 2007; 190:255–9. [PubMed: 17329747]
5. Sollid CP, Wisborg K, Hjort J, Secher NJ. Eating disorder that was diagnosed before pregnancy and pregnancy outcome. *American Journal of Obstetrics and Gynecology*. 2004; 190:206–10. [PubMed: 14749661]
6. Bulik CM, Sullivan PF, Fear JL, Pickering A, Dawn A, McCullin M. Fertility and reproduction in women with anorexia nervosa: A controlled study. *Journal of Clinical Psychiatry*. 1999; 60:130–5. quiz 5-7. [PubMed: 10084645]
7. Reba-Harrelson L, Von Holle A, Hamer RM, Torgersen L, Reichborn-Kjennerud T, Bulik CM. Patterns of maternal feeding and child eating associated with eating disorders in the Norwegian Mother and Child Cohort Study (MoBa). *Eating Behaviors*. 2010; 11:54–61. [PubMed: 19962121]
8. Watson HJ, Von Holle A, Hamer RM, Knoph Berg C, Torgersen L, Magnus P, et al. Remission, continuation and incidence of eating disorders during early pregnancy: A validation study in a population-based birth cohort. *Psychological Medicine*. 2012:1–12.
9. Lubetzky-Vilnai A, Ciol M, McCoy SW. Statistical analysis of clinical prediction rules for rehabilitation interventions: current state of the literature. *Archives of Physical Medicine and Rehabilitation*. 2014; 95(1):188–96. [PubMed: 24036159]
10. Steyerberg, EW. *Clinical prediction models: A practical approach to development, validation, and updating (statistics for biology and health)*. Springer; New York: 2009.
11. Magnus P, Irgens L, Haug K, Nystad W, Skjaerven R, Stoltenberg C. Cohort profile: The Norwegian Mother and Child Cohort Study (MoBa). *International Journal of Epidemiology*. 2006; 35:1146–50. [PubMed: 16926217]
12. American Psychiatric Association. *Diagnostic and statistical manual of mental disorders*. 4th ed. American Psychiatric Association; Washington, DC: 2000.
13. Bulik CM, Von Holle A, Hamer R, Knoph Berg C, Torgersen L, Magnus P, et al. Patterns of remission, continuation and incidence of broadly defined eating disorders during early pregnancy in the Norwegian Mother and Child Cohort Study (MoBa). *Psychological Medicine*. 2007; 37(8): 1109–18. [PubMed: 17493296]
14. Reichborn-Kjennerud T, Bulik CM, Tambs K, Harris JR. Genetic and environmental influences on binge eating in the absence of compensatory behaviors: A population-based twin study. *International Journal of Eating Disorders*. 2004; 36(3):307–14. [PubMed: 15478129]
15. Tambs K, Moum T. How well can a few questionnaire items indicate anxiety and depression? *Acta Psychiatrica Scandinavica*. 1993; 87:364–7. [PubMed: 8517178]
16. Strand BH, Dalgard OS, Tambs K, Rognerud M. Measuring the mental health status of the Norwegian population: A comparison of the instruments SCL-25, SCL-10, SCL-5 and MHI-5 (SF-36). *Nordic Journal of Psychiatry*. 2003; 57:113–8. [PubMed: 12745773]
17. Diener E, Emmons RA, Larsen RJ, Griffin S. The Satisfaction With Life Scale. *Journal of Personality Assessment*. 1985; 49(1):71–5. [PubMed: 16367493]
18. Pavot W, Diener E. Review of the Satisfaction With Life Scale. *Psychological Assessment*. 1993; 5:164–72.
19. Rosenberg, M. *Society and the adolescent self-image*. Wesleyan University Press; Middletown, CT: 1989.
20. Tambs K. Moderate effects of hearing loss on mental health and subjective well-being: results from the Nord-Trøndelag Hearing Loss Study. *Psychosomatic Medicine*. 2004; 66(5):776–82. [PubMed: 15385706]
21. Røysamb, E.; Vittersø, J.; Tambs, K. *The Relationship Satisfaction Scale: Reliability, validity and goodness of fit*. unpublished

22. Røsand GM, Slinning K, Eberhard-Gran M, Røysamb E, Tambs K. Partner relationship satisfaction and maternal emotional distress in early pregnancy. *BMC Public Health*. 2011; 11:161. [PubMed: 21401914]
23. Blum JS, Mehrabian A. Personality and temperament correlates of marital satisfaction. *Journal of Personality*. 1999; 67:93–125.
24. Norton R. Measuring marriage quality: A critical look at the dependent variable. *Journal of Marriage and the Family*. 1983; 45:141–51.
25. Kendler KS, Neale MC, Kessler RC, Heath AC, Eaves LJ. The lifetime history of major depression in women: Reliability of diagnosis and heritability. *Archives of General Psychiatry*. 1993; 50:863–70. [PubMed: 8215812]
26. Grobbee, DE.; Hoes, AW. *Clinical epidemiology: Principles, methods, and applications for clinical research*: Jones & Bartlett Learning. 2009.
27. Watson H, Von Holle A, Hamer R, Knoph Berg C, Torgersen L, Magnus P, et al. Remission, continuation and incidence of eating disorders during early pregnancy: A validation study in a population-based birth cohort. *Psychological Medicine*. 2013; 43(8):1723–34. [PubMed: 23164164]
28. Efron, B.; Tibshirani, R. *An introduction to the bootstrap*. Chapman and Hall; London: 1993.
29. Good, PI. *Resampling methods: A practical guide to data analysis*. Birkhäuser; Boston: 2006.
30. Benjamini Y, Hochberg Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society Series B (Methodological)*. 1995:289–300.
31. Steyerberg EW, Borsboom GJJM, van Houwelingen HC, Eijkemans MJC, Habbema JDF. Validation and updating of predictive logistic regression models: A study on sample size and shrinkage. *Statistics in Medicine*. 2004; 23(16):2567–86. [PubMed: 15287085]
32. Harrell FE, Lee KL, Califf RM, Pryor DB, Rosati RA. Regression modelling strategies for improved prognostic prediction. *Statistics in Medicine*. 1984; 3(2):143–52. [PubMed: 6463451]
33. Nilsen RM, Vollset SE, Gjessing HK, Skjærven R, Melve KK, Schreuder P, et al. Self-selection and bias in a large prospective pregnancy cohort in Norway. *Paediatric and Perinatal Epidemiology*. 2009; 23(6):597–608. [PubMed: 19840297]
34. Nilsen RM, Surén P, Gunnes N, Alsaker ER, Bresnahan M, Hirtz D, et al. Analysis of Self-selection Bias in a Population-based Cohort Study of Autism Spectrum Disorders. *Paediatric and Perinatal Epidemiology*. 2013; 27(6):553–63. [PubMed: 23919580]
35. Berg KC, Peterson CB, Frazier P, Crow SJ. Psychometric evaluation of the Eating Disorder Examination and Eating Disorder Examination-Questionnaire: A systematic review of the literature. *International Journal of Eating Disorders*. 2012; 45:428–38. [PubMed: 21744375]
36. Kvalvik LG, Nilsen RM, Skjærven R, Vollset SE, Midttun O, Ueland PM, Haug K. Self-reported smoking status and plasma cotinine concentrations among pregnant women in the Norwegian Mother and Child Cohort Study. *Pediatric Research*. 2012; 72:101–7. [PubMed: 22441375]

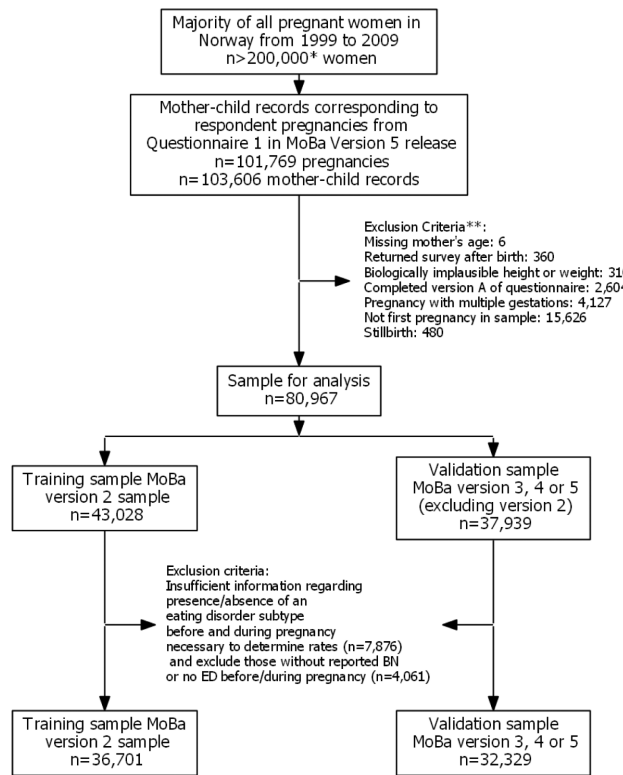


Figure 1. Participant flow to achieve final analysis sample. MoBa, The Norwegian Mother and Child Cohort Study. *Extrapolated from the reported 38.5% participation rate (<http://www.fhi.no>). **Criteria not mutually exclusive

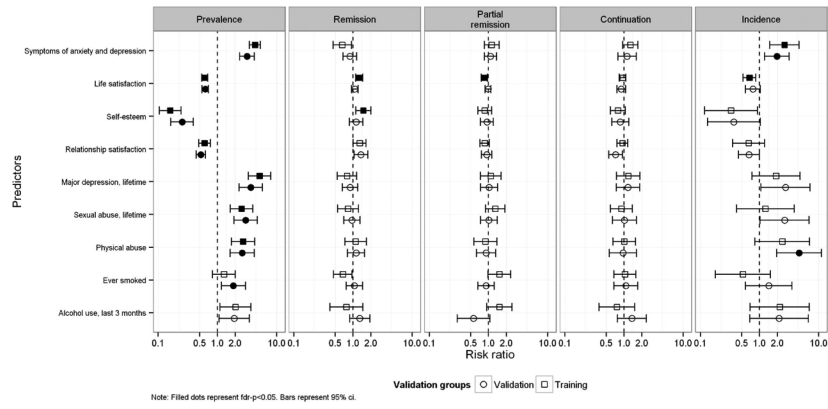


Figure 2. Rate ratios and 95% confidence intervals (CIs) for the association between psychosocial factors and BN outcomes by sample.

Table 1
 Psychosocial characteristics of women in the MoBa cohort in the total sample (N = 69,030)

Characteristic	Outcome			
	Remission (<i>n</i> = 195, 0.3%)	Partial remission (<i>n</i> = 182, 0.3%)	Continuation (<i>n</i> = 132, 0.2%)	No ED during pregnancy (<i>n</i> = 68,478, 99.2%)
Symptoms of anxiety and depression, <i>M</i> (<i>SD</i>)	1.60 (0.62)	1.71 (0.64)	1.76 (0.73)	1.49 (0.49)
Life satisfaction, <i>M</i> (<i>SD</i>)	5.10 (1.30)	4.73 (1.40)	4.65 (1.55)	5.20 (1.20)
Self-esteem, <i>M</i> (<i>SD</i>)	2.93 (0.65)	2.82 (0.62)	2.74 (0.78)	3.03 (0.68)
Relationship satisfaction, <i>M</i> (<i>SD</i>)	5.25 (0.67)	5.05 (0.92)	4.91 (0.84)	5.13 (0.80)
Lifetime major depression, <i>n</i> (%)	101 (51.8)	103 (56.6)	77 (58.3)	19 (44.2)
Sexual abuse, <i>n</i> (%)	71 (36.4)	75 (41.2)	50 (37.9)	13 (30.2)
Physical abuse, <i>n</i> (%)	59 (30.3)	49 (26.9)	39 (29.5)	14 (32.6)
Ever smoked, <i>n</i> (%)	113 (57.8)	121 (66.5)	86 (65.2)	20 (46.5)
Alcohol use twice p/wk, <i>n</i> (%)	26 (13.3)	25 (13.7)	18 (13.6)	10 (23.3)

ED, eating disorder.

Table 2

Bootstrap risk ratio and bias estimates in total sample by psychosocial characteristics

	Bootstrap (95% CI) Bias (difference in regression coefficients) (se) Original (95% CI)				
	Prevalence	Remission	Partial remission	Continuation	Incidence
Symptoms of anxiety and depression	3.72 (3.10 - 4.42) -0.01 (0.09)	0.78 (0.61 - 0.97) -0.01 (0.12)	1.10 (0.92 - 1.32) 0.00 (0.09)	1.23 (0.97 - 1.55) -0.00 (0.12)	2.28 (1.50 - 3.35) -0.03 (0.20)
Life satisfaction	3.72 (3.14 - 4.41) 0.62 (0.56 - 0.67)	0.78 (0.63 - 0.97) 1.16 (1.05 - 1.28)	1.10 (0.92 - 1.31) 0.92 (0.85 - 1.00)	1.23 (0.98 - 1.54) 0.92 (0.83 - 1.02)	2.31 (1.57 - 3.38) 0.73 (0.59 - 0.88)
Self-esteem	0.00 (0.05) 0.20 (0.15 - 0.27)	0.00 (0.05) 1.16 (1.05 - 1.28)	0.00 (0.04) 0.92 (0.85 - 1.00)	0.00 (0.05) 0.92 (0.83 - 1.02)	0.01 (0.10) 0.37 (0.17 - 0.73)
Relationship satisfaction	0.01 (0.16) 0.20 (0.15 - 0.27)	0.00 (0.10) 1.28 (1.05 - 1.56)	-0.00 (0.09) 0.91 (0.76 - 1.09)	0.00 (0.11) 0.81 (0.65 - 1.01)	0.01 (0.37) 0.35 (0.17 - 0.71)
Lifetime major depression	0.57 (0.49 - 0.66) 0.00 (0.07)	1.35 (1.12 - 1.61) 0.01 (0.09)	0.91 (0.79 - 1.04) 0.01 (0.07)	0.83 (0.70 - 0.98) -0.01 (0.09)	0.68 (0.44 - 1.01) 0.02 (0.21)
Sexual abuse	0.00 (0.16) 4.37 (3.19 - 5.97)	-0.00 (0.12) 1.34 (1.11 - 1.60)	0.00 (0.13) 1.05 (0.81 - 1.35)	0.00 (0.17) 1.21 (0.88 - 1.67)	-0.02 (0.35) 2.29 (1.18 - 4.45)
Physical abuse	2.79 (1.98 - 3.81) -0.00 (0.17)	0.90 (0.70 - 1.15) -0.00 (0.13)	1.14 (0.88 - 1.46) 0.00 (0.13)	0.97 (0.69 - 1.32) -0.00 (0.17)	1.99 (0.89 - 3.91) -0.01 (0.38)
Ever smoked	2.75 (2.01 - 3.77) 2.66 (1.87 - 3.66)	0.90 (0.70 - 1.15) 1.10 (0.83 - 1.41)	1.13 (0.89 - 1.45) 0.92 (0.67 - 1.23)	0.96 (0.70 - 1.31) 1.02 (0.71 - 1.42)	1.89 (0.92 - 3.86) 3.52 (1.61 - 6.74)
Alcohol use 2 p/wk	-0.02 (0.17) 2.66 (1.92 - 3.69)	-0.00 (0.13) 1.09 (0.84 - 1.41)	-0.01 (0.15) 0.91 (0.69 - 1.22)	-0.01 (0.18) 1.01 (0.73 - 1.41)	-0.03 (0.37) 3.40 (1.74 - 6.66)
	1.54 (1.10 - 2.08) -0.01 (0.16)	0.87 (0.68 - 1.09) 0.00 (0.12)	1.14 (0.86 - 1.47) 0.01 (0.13)	1.07 (0.77 - 1.45) -0.00 (0.16)	0.92 (0.43 - 1.73) -0.01 (0.35)
	1.53 (1.11 - 2.10) 1.98 (1.27 - 2.96)	0.86 (0.68 - 1.08) 1.11 (0.76 - 1.56)	1.12 (0.87 - 1.44) 0.89 (0.58 - 1.31)	1.06 (0.78 - 1.45) 1.04 (0.65 - 1.58)	0.87 (0.45 - 1.70) 2.26 (0.87 - 4.91)
	-0.01 (0.22) 1.95 (1.28 - 2.98)	-0.02 (0.18) 1.11 (0.79 - 1.56)	-0.00 (0.21) 0.88 (0.59 - 1.31)	-0.02 (0.23) 1.03 (0.67 - 1.60)	-0.04 (0.44) 2.14 (0.94 - 4.86)

All FDR adjusted *p*-values significant at alpha level of 0.05 are in bold indicating a risk ratio significantly different from 1.

Values are adjusted for maternal age and education.

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Table 3

Overall beta estimates (se) from calibration models

Characteristic	Outcome				
	Prevalence	Remission	Partial remission	Continuation	Incidence
Symptoms of anxiety and depression	0.82 (0.1)	0.30 (0.2)	0.13 (0.6)	0.90 (0.4)	0.33 (0.2)
Life satisfaction	0.92 (0.1)	0.26 (0.2)	-0.18 (0.3)	0.46 (0.4)	0.22 (0.2)
Self-esteem	0.77 (0.1)	0.22 (0.2)	-0.38 (0.6)	0.69 (0.4)	0.44 (0.2)
Relationship satisfaction	1.00 (0.2)	0.67 (0.3)	0.11 (0.6)	1.08 (0.5)	0.25 (0.2)
Lifetime major depression	0.73 (0.1)	0.28 (0.2)	-0.60 (0.9)	0.51 (0.4)	0.35 (0.2)
Sexual abuse	0.88 (0.2)	0.29 (0.3)	-0.12 (0.5)	0.52 (0.4)	0.18 (0.2)
Physical abuse	0.69 (0.2)	0.14 (0.3)	-0.19 (0.7)	0.49 (0.5)	0.38 (0.2)
Ever smoked	0.46 (0.3)	0.21 (0.3)	-0.26 (0.3)	0.44 (0.5)	0.04 (0.3)
Alcohol use twice p/wk	0.57 (0.3)	0.18 (0.3)	-0.45 (0.5)	0.16 (0.5)	0.08 (0.2)

All FDR adjusted *p*-values significant at alpha level of 0.05 are in bold indicating a β significantly different from 1.

Values are adjusted for maternal age and education.

Table 4

Brier scores for the calibration models

Characteristic	Outcome				
	Prevalence	Remission	Partial remission	Continuation	Incidence
Symptoms of anxiety and depression	0.0024	0.5540	0.2363	0.1761	0.0006
Life satisfaction	0.0025	0.5726	0.2684	0.1717	0.0006
Self-esteem	0.0025	0.4846	0.2499	0.1689	0.0007
Relationship satisfaction	0.0025	0.2594	0.2425	0.1665	0.0007
Lifetime major depression	0.0025	0.3186	0.2370	0.1709	0.0007
Sexual abuse	0.0025	0.2668	0.2385	0.1895	0.0007
Physical abuse	0.0025	0.2522	0.2357	0.1698	0.0007
Ever smoked	0.0026	0.3245	0.2452	0.1729	0.0007
Alcohol use > twice p/wk	0.0025	0.2683	0.2446	0.1704	0.0007