British Journal of Psychiatry (1989), **155**, 356–362

Maternity Blues

I. Detection and Measurement by Questionnaire

HELEN KENNERLEY and DENNIS GATH

The studies reported had two purposes: (a) to identify psychological symptoms commonly experienced by mothers in the first few days after childbirth; (b) to use psychometric methods to develop a questionnaire for detecting and measuring these symptoms. A draft 49-item questionnaire was derived by asking 100 newly delivered mothers to volunteer suitable items. The number of items was systematically reduced to 28 while testing the questionnaire on two further samples of newly delivered mothers ($n = 100$ and $n = 50$). Cluster analysis of the 28 items yielded seven clusters, the robustness of which was established in a further sample of 87 newly delivered mothers. The validity of the 28-item questionnaire was established. The questionnaire was quick and easy to administer, suitable for repeated administration on consecutive days, and well accepted by mothers.

'Maternity blues' usually denotes a brief psychological disturbance in mothers in the first few days after childbirth. Maternity blues is generally regarded as a disturbance of mood, but there is little agreement on its precise definition. Weeping and lability of mood are commonly described, but there is disagreement as to whether or not depressed mood is characteristic (Yalom *et al*, 1968; Handley *et al*, 1980; Cox *et al*, 1982). It is usually said that symptoms reach a peak of intensity about four or five days after birth, and then decline fairly rapidly, disappearing before the tenth day. Estimates of the frequency of the blues have ranged widely from 15% of births (Oppenheim, 1962) to 80% (Robin, 1962). More information about research on the blues can be found in reviews by Stein (1982) and by Kennerley & Gath (1986).

Maternity blues is brief and benign, and therefore not a serious problem in clinical practice. Nevertheless, it has interested researchers because it might throw light on the aetiology of mood disorders. So far, most research findings have been inconclusive, possibly because instruments for detecting and measuring the blues have varied in their suitability. For example, standardised psychiatric interviews, such as the Present State Examination (Wing *et al*, 1974) and the Standardised Psychiatric Interview (Goldberg *et al*, 1970), have been used to measure psychiatric symptoms in the early puerperium (George *et al*, 1980; Cox *et al*, 1982). Although well established, these instruments were not designed for measuring the blues, and are not suitable for day-to-day use.

As mood changes may be rapid during the early puerperium, daily assessments are essential. For this purpose a self-evaluation schedule is more practical than daily interviews. Some researchers have studied the blues with self-rating scales originally designed to measure depressive symptoms (Nott *et al*, 1976; Handley *et al*, 1980; Ballinger *et al*, 1982; Manley *et al*, 1982; Cutrona, 1983). Such measures may be better for day-to-day use, but may still be inappropriate because it is not established that depressive symptoms are the main feature of the blues.

Three scales have been designed for measuring the blues (Pitt, 1973; Stein, 1980; Kendell *et al*, 1981), although not developed by systematic psychometric methods. These three scales have between them a total of 28 different items, of which only three are common to all three.

For these reasons, it seemed desirable to devise a further questionnaire to detect and measure maternity blues. The aims of the studies described in this paper were:

(a) to identify psychological symptoms commonly experienced by women in the ten days after childbirth
(b) to use psychometric methods to develop a questionnaire which is suitable for detecting and measuring maternity blues by repeated administration on consecutive days, is short and easy to administer, and is acceptable to mothers
(c) to determine the validity of this questionnaire.

Procedures

Stage 1: compilation of possible items for first draft questionnaire

One hundred newly delivered mothers (group A) were selected for interview in the maternity department of the

https://doi.org/10.1192/bjp.155.3.356 Published online by Cambridge University Press

TABLE I
Stages of elimination and modification of questionnaire items

Item no.	*Original item (Stage 1)*	*Response frequency (%)*	*Stage 2*		*Stage 3*	
			Redundant	*Infrequent or non-specific*	*Non-peaking*	*Items as they appear in final questionnaire*
1	Tearful	47				Tearful
2	Happy	40				Happy
3	Concentrating poorly	40				Able to concentrate
4	Lively	37				Lively
5	Confident	36				Confident
6	Weeping	33	Similar to 1			
7	Tolerant	33		✓		
8	Calm	32				Calm
9	Tired	32				Tired
10	Elated	30				Elated
11	Sleepy	30	Similar to 9			
12	Physically fit	26		✓		
13	Alert	24				Alert
14	Oversensitive	24				Oversensitive
15	Restless	23				Restless
16	Overemotional	22				Overemotional
17	Muddled thinking	21	Similar to 21/41			
18	Crying, unable to stop	20				Crying without being able to stop
19	Excited	19	Similar to 10/30			
20	Low spirited	17				Low spirited
21	Forgetful	17				Forgetful, muddled
22	Mentally relaxed	16				Mentally relaxed
23	Physically relaxed	16		✓		
24	Emotionally reserved	13				Finding it difficult to show my feelings
25	Relieved	11		✓		
26	Brooding	11				Brooding on things
27	Optimistic	11		✓		
28	Purposeful	11			✓	
29	Emotionally stable	11				Up and down in mood
30	Content	10	Similar to 10/19			
31	Over-reacting	10	Similar to 14			Changeable in mood
32	Changeable in mood	9				
33	Bonding poorly	9			✓	
34	Mentally tense	7				Mentally tense
35	Lacking interest	6			✓	
36	Care-free	5	Similar to 22			
37	Irritable	5				Irritable
38	Hostile, touchy	5	Similar to 37			
39	Emotionally numb	4				Emotionally numb, without feelings
40	Physically tense	4		✓		
41	Slow in thinking	3	Similar to 17/21			
42	Self-pitying	2				Feeling sorry for myself
43	Helpless	2				Helpless
44	Socially withdrawn	2				Wishing that I was alone
45	Depressed	2				Depressed
46	Disorientated	2		✓		
47	Anxious	1				Anxious
48	Earnest	1		✓		
49	Possessive	1			✓	

https://doi.org/10.1192/bjp.155.3.356 Published online by Cambridge University Press

local teaching hospital. Each mother was interviewed once in the ten days immediately after childbirth: 74 interviews were held on days 1–6; and 26 on days 7–10, most mothers having left hospital by then. This difference was not thought important because the purpose was simply to compile suitable adjectives for the first draft questionnaire.

Interviewing was unstructured. Mothers were simply asked to describe in their own words the feelings they had experienced at any time that day, and to indicate whether these feelings differed in any way from their normal feelings (before pregnancy). Most mothers had no difficulty in putting their feelings into words. The interviewer helped the few who had difficulty by suggesting some common mood adjectives to start them talking. Interviews lasted about 30 minutes and all were tape-recorded.

Analysis of items

Of the 100 women interviewed in group A, 47 reported unusual feelings on the day in question; all 47 were interviewed on one of the first five days. Of the 53 women reporting no such experiences, 21 were interviewed in the first five days, and 32 on days 6–10.

The tape-recordings were used to compile a list of adjectives used by the 47 women, and to determine their frequency (Table I). It is notable that the term 'depressed' was used by only 2% of respondents, while 'low spirited' was used by 17%.

'Confusion' has often been cited as a feature of maternity blues. In the present sample, 'concentrating poorly' and 'muddled thinking' were commonly reported. However, there was no evidence that these symptoms were associated with organic mental states, since item 46, 'disorientated', was reported by only 2% of the mothers, and at interview there was no clinical evidence of either disorientation or clouding of consciousness in any respondent.

Stage 2: administration and revision of first draft questionnaire

The 49 items elicited in stage 1 were assembled in random order in a draft questionnaire. Each item was rated on the following five-point scale: much less than usual; less than usual; same as usual; more than usual; much more than usual.

The questionnaire was administered to a second group of 100 mothers (group B) who were in the maternity ward for up to ten days after childbirth. It was then shortened by eliminating items which, as shown in Table I, were:

(a) redundant, that is, highly correlating items which appeared to be measuring the same psychological dimension (items 6, 11, 17, 19, 30, 31, 36, 38, 41)
(b) infrequently endorsed, that is, rated by less than 15% of mothers (items 7, 46, 48)
(c) non-specific (items 25, 27) or descriptive of physical states (items 12, 23, 40).

Stage 3: administration and revision of second draft questionnaire

The second draft questionnaire (as revised in stage 2) was given to a further 50 mothers (group C) who were in hospital up to ten days after childbirth. Each mother was asked to complete the questionnaire on the day it was given to her and again on the following day. Ratings were to be made in the evening, and were to apply to the day as a whole.

At this stage four items were removed, as they failed to show a peak of frequency in the first few days (Table I, items 28, 33, 35, 49). These were omitted because past research findings and clinical observations had suggested that peaking is characteristic of the blues. In this way the number of items was reduced to its final total of 28. The final version of the questionnaire is shown in the Appendix.

Stage 4: cluster analysis and scoring system

Cluster analysis and labelling of clusters

The next step was to determine whether the 28-item questionnaire was measuring one or more different syndromes. For this purpose, the statistical technique of cluster analysis was preferred to other forms of factor analysis because it has been specifically designed as a means of sorting sample data in terms of typography. This technique is especially appropriate when little is known about category structure within the data. The purpose of the technique is to sort data into clusters such that the degree of association is high among members of the same cluster, and low among members of different clusters (Everitt, 1974). In our analysis, mothers' responses were subjected to Ward's method of clustering, using the computer program CLUSTAN (Wishart, 1975).

The clusters comprised items, not subjects, and the similarity measure used was based on the squared Euclidean distances between items. Analysis using Ward's method of clustering produced a dendogram of clusters. The cut-off point used to determine the most meaningful grouping of items is arbitrary, but it is generally taken to be the first large increase in increment in the error coefficient (the latter being the sum of distances from each individual item to the centroid of its parent cluster).

To ensure that the scale would measure change from normal state, and that clusters would be homogeneous, responses were entered into the cluster analysis only when the respondent rated seven or more of the 28 items as applying to her on the day in question. Of the 100 sets of responses (50 group C subjects × 2 days), the number meeting this criterion was 32. Seven clusters emerged (Table II). From inspection of the components of clusters, and bearing in mind the direction of rating (see 'scoring system', below), the investigators gave a label to each cluster, as shown in Table II.

Scoring system

When the daily frequencies of the 28 items were plotted graphically (not shown) it was found that 20 items showed a peak when rated 'more than usual', and eight items when rated 'less than usual' (Table II).

It was decided to allocate a score of 1 to each of the 20 items showing a peak when rated 'more than usual' or 'much more than usual', and to each of the eight items showing a peak when rated 'less than usual', or 'much less

https://doi.org/10.1192/bjp.155.3.356 Published online by Cambridge University Press

TABLE II
Composition and labelling of clusters

Cluster label	*Component items*	
	Group C (n = 32)	*Group D (n = 87)*
1. Primary blues	Tearful	Tearful
	Tired	—
	Anxious	—
	Overemotional	Overemotional
	Changeable in mood	Changeable in mood
	Low spirited	Low spirited
	Forgetful/muddled	—
		Up and down in mood
		Oversensitive
2. Reservation	Emotionally reserved	Emotionally reserved
	Emotionally numb without feeling	Emotionally numb without feeling
	Socially withdrawn	Socially withdrawn
3. Hypersensitivity	Mentally tense	
	Restless	
	Oversensitive	
	Up and down in mood	
4. Depression	Brooding on things	Brooding on things
	Irritable	Irritable
	Depressed	Depressed
	Self-pitying	Self-pitying
	Crying without being able to stop	—
		Confident[1]
		Mentally tense
		Anxious
		Restless
5. Despondency	Elated[1]	Elated[1]
	Helpless	—
	Happy[1]	Happy[1]
6. Retardation	Able to concentrate[1]	Able to concentrate[1]
	Alert[1]	Alert[1]
	Lively[1]	Mentally relaxed[1]
		Calm[1]
7. Decreased self-confidence	Confident[1]	
	Mentally relaxed[1]	
	Calm[1]	
New cluster 1		Tired[1]
		Lively
New cluster 2		Helpless
		Forgetful/muddled
		Crying without being able to stop

1. Items rated in direction 'less than usual' (all other items rated 'more than usual').

than usual'. In this way, each respondent could be given an overall score (maximum 28) and a subscore for each cluster (the maximum varying according to the number of items in the cluster), thereby indicating the nature and the severity of her condition.

Thus we used the data categorically (absent = 0; present = 1) and did not attempt to measure severity by using the data dimensionally. This was because a range of responses might not indicate severity but rather individual scoring style (i.e. tendency to be conservative or extreme). In paper III we describe how this system can be employed in the analysis of continuous and discrete variables.

If preferred, the five-point scale (see Appendix) could be used to assess severity of an item by scoring 'more (less) than usual' as 1 and 'much more (less) than usual' as 2. We suggest that the number of options should not be increased to more than these five as it has been established that more complex scoring methods confer no real advantage (Likert, 1932).

Cluster frequency

By using the categorical scoring system, it was possible to determine the frequency of each cluster among the 50

https://doi.org/10.1192/bjp.155.3.356 Published online by Cambridge University Press

women in group C. This was done for each cluster by determining how many women scored on more than half the items in the cluster. On this basis the proportions of respondents were: primary blues 36%; retardation 36%; hypersensitivity 26%; decreased self-confidence 24%; depression 16%; reservation 10%; despondency 8%.

Stage 5: administration of final 28-item questionnaire

Testing the robustness of clusters

The final 28-item questionnaire was completed by a further sample of 87 newly delivered women (group D) every day during the first ten days after childbirth.

By comparing group C and group D, the robustness of clusters was shown in two ways. (a) When the responses of group D were subdivided according to the clusters originally derived from group C responses, the frequency of the seven clusters followed almost the same rank order in both groups. (b) When a new cluster analysis was carried out on group D responses, the new clusters were similar in composition to the original group C clusters. This degree of resemblance is unusually close considering that cluster analysis was applied to a different group at a different time.

Day-to-day variation in responses

Figure 1 shows the mean percentage scores on the questionnaire for group D women on each of the ten post-natal days. Mean percentage scores were derived as follows. For each individual woman, a total score for each of the ten days was expressed as a percentage of the items completed by her (this was done to allow for occasional missing values). A mean percentage score for each of the ten days was then derived for the sample as a whole, together with an overall mean value for the ten days as a whole. It can be seen that the scores reached a peak (24.3%) on day 5, and tailed off towards day 10. Analysis of variance, however, showed that the pattern of distribution of the scores did not differ significantly from chance. (However, a subsequent study using this questionnaire has shown statistically significant peaking of scores – see 'Discussion'.)

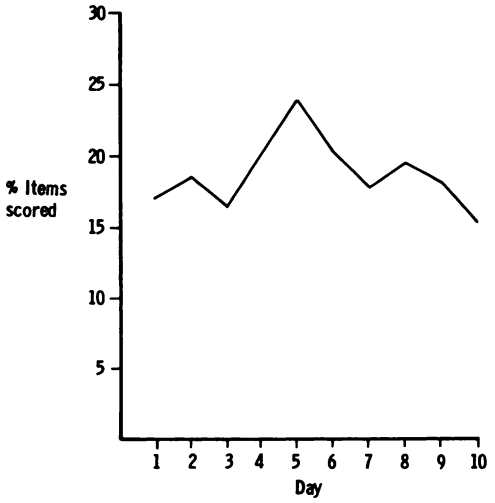


FIG. 1 Mean scores on the questionnaire during the first ten days after childbirth. For each woman, a total score for each of the ten days was expressed as a percentage of the items completed by her (to allow for occasional missing values). A mean percentage score for each of the ten days was then derived for the sample as a whole.

Figure 2 shows the percentage item scores for individual clusters for group D women on each of the ten post-natal days. It can be seen that the clusters appear to co-vary, and most clusters tend to peak in frequency about the fifth day.

Validity of the questionnaire

The questionnaire was devised to detect symptoms which might change rapidly. For this reason the reliability of the questionnaire could not be tested by administering it to the same woman on separate occasions. However, when the questionnaire was given first to group C and then to the similar but completely independent sample, group D, response patterns were found to be stable in two ways: first, symptoms tended to peak on the same post-partum day in both groups; second, as explained above, patterns of clustering were very similar in the two groups.

Content validity was supported in three ways: the use of open-ended inquiry to elicit an original pool of items from 100 mothers (group A) is one of the most powerful demonstrations of content validity possible; the samples of mothers were highly representative because over 95% of mothers from the catchment area are delivered in the hospital concerned; and four different samples of mothers were used.

Face validity was supported by the fact that subsequent groups of mothers found the questionnaire items acceptable and appropriate.

Construct validity was shown in the following four ways.

(a) 'Primary blues' emerged as a distinct cluster from 'depression' and was much more frequent. This was true for mothers in both groups C and D.

(b) Of the 87 mothers in group D, 79 completed the questionnaire not only post-natally but also over a three-day period four weeks antenatally. Scores were significantly higher post-natally. Thus the mean total percentage of items endorsed increased from 14.8% antenatally to 21.3% post-natally ($t=2.85$; d.f. 78; $P<0.001$), while the mean peak score increased from 23.6 to 45.7 ($t=6.06$; d.f. 78; $P<0.001$). It is important to note that cluster scores changed significantly in only two clusters. Thus the score for 'primary blues' increased from 19.6 antenatally to 27.8 post-natally ($t=3.78$; d.f. = 11; $P<0.01$), and for 'hypersensitivity' from 17.6 antenatally to 22.6 post-natally ($t=3.45$; d.f. = 1; $P<0.01$). Scores for the remaining clusters (which contained some items traditionally ascribed to 'maternity blues') did not change. These findings supported the argument that changes in depression are not characteristic of the early puerperium.

(c) All clusters except despondency showed a tendency to peak about day 5 (in line with symptom peaking as reported by other investigators).

https://doi.org/10.1192/bjp.155.3.356 Published online by Cambridge University Press

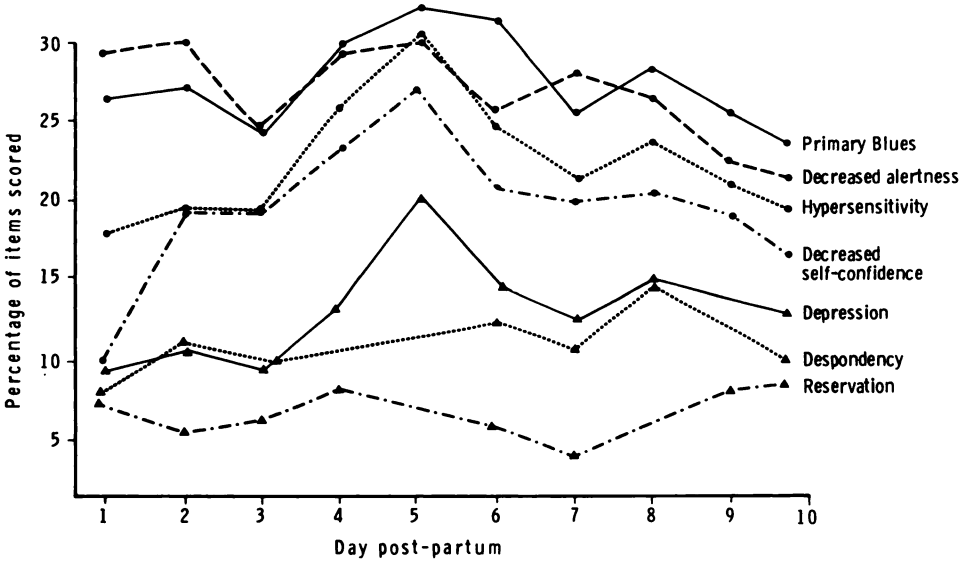


FIG. 2 Cluster scores on the questionnaire during the first ten days after childbirth.

(d) In a separate study (paper II) it was found that whereas questionnaire scores tended to peak on the fourth or fifth post-natal day in maternity patients, in surgical patients they tended to peak on the second post-operative day and then declined steadily.

Discussion

Asking subjects to volunteer words describing their feelings gives the advantage that items elicited are both relevant and acceptable to the population studied (Mackay, 1980). By eliminating redundant items by systematic item analysis, the final questionnaire was reduced to 28 items which could be grouped into seven clusters. Studying a further sample confirmed the validity of the questionnaire and established the robustness of clusters.

Our questionnaire differs in several ways from the three best-known measures designed to rate the blues. Among the latter, the rating scale devised by Pitt (1973) was based on Hamilton's (1962) account of the blues, which listed eight symptoms that midwives regarded as most frequent in the mid-puerperium: fatigue, crying, anxiety, confusion, headache, insomnia, hypochondriasis, and hostility to husband. Pitt added 'depression' to this list. In our questionnaire, there are only four items that appear to correspond to these nine items (tired, tearful, anxious, and, possibly, muddled thinking).

The questionnaire devised by Stein (1980) contained 13 items (depression, weeping, restlessness, anxiety, fatigue, tension, irritability, poor concentration, forgetfulness, confusion, anorexia, dreaming, and headache). The first nine of these items correspond approximately to items in our questionnaire. Another of Stein's items, 'confusion', may correspond to our 'forgetful/muddled'. Three of Stein's items (anorexia, dreaming, and headaches) have no counterpart in our questionnaire, and 18 of our items do not appear in Stein's questionnaire. It is notable that all these differences between the two questionnaires were present in stage 1 of the development of our questionnaire. In other words the differences may have occurred because our first procedure was based on eliciting adjectives directly from the mothers.

Our questionnaire contained corresponding items for all six items which Kendell *et al* (1981) included in their six visual analogue scales, namely happiness, depression, tears, anxiety, irritability, and lability.

Two points about our questionnaire are stressed. First, when the questionnaire was administered over a ten-day period, mean scores appeared to peak on day 5, but day-to-day variation was not statistically significant. Subsequently, however, Owen & Campbell (1987) gave the questionnaire to a further sample of women and found a statistically significant peaking of scores on the third day post-partum.

Second, the clusters threw some light on the question of whether depression is characteristic of maternity blues. The most frequent cluster, which we called 'primary blues', included the item 'low spirited', whereas a less frequent cluster, 'depression',

https://doi.org/10.1192/bjp.155.3.356 Published online by Cambridge University Press

included the item 'depressed'. A possible explanation is that the mothers used 'low spirited' to denote a demoralised state of low intensity, and 'depressed' to indicate a more severe dysphoria. Mothers frequently stated, "I don't feel depressed but I do feel low spirited". Thus the findings suggested that depression is not typical of the blues.

Conclusions

Our questionnaire is valid for detecting and measuring specific brief psychological disturbances occurring in the first few days after childbirth. The questionnaire should be suitable for future research on the aetiology of maternity blues (whether psychological or biological) and on the associations between the blues and previous or subsequent psychiatric disorder.

Appendix

The questionnaire is reproduced below. It was modified to include the columns headed 'no' and 'yes' because women said that it would be easier to answer the questionnaire in two steps – deciding first whether the item applied to them at all, and then (if it did apply) deciding to what extent. This change made the instrument more acceptable to mothers without altering the psychometric properties of the data collected.

BLUES QUESTIONNAIRE : Kennerley

Subject Name: No: Date: Days Postpartum:

Below is a list of words which newly delivered mothers have use to describe how they are feeling. Please indicate **HOW YOU HAVE BEEN FEELING TODAY** by ticking NO or YES. Then please mark the box which best describes how much change there is, if any, from your usual self.

	No	Yes	IS THIS	Much less than usual	Less than usual	No different	More than usual	Much more than usual
1. Tearful								
2. Mentally tense								
3. Able to concentrate								
4. Low spirited								
5. Elated								
6. Helpless								
7. Finding it difficult to show your feelings								
8. Alert								
9. Forgetful, muddled								
10. Anxious								
11. Wishing you were alone								
12. Mentally relaxed								
13. Brooding on things								
14. Feeling sorry for yourself								
15. Emotionally numb, without feelings								
16. Depressed								
17. Over-emotional								
18. Happy								
19. Confident								
20. Changeable in your spirits								
21. Tired								
22. Irritable								
23. Crying without being able to stop								
24. Lively								
25. Over-sensitive								
26. Up and down in your mood								
27. Restless								
28. Calm, tranquil								

Thank you for your co-operation

References and authors' details will be found at the end of paper III, on page 373.

https://doi.org/10.1192/bjp.155.3.356 Published online by Cambridge University Press