

# Field trials of the Baby Check score card in hospital

A J Thornton, C J Morley, T J Cole, S J Green, K A Walker, J M Rennie

## Abstract

The Baby Check score card was used by junior paediatric doctors to assess 262 babies under 6 months old presenting to hospital. The duty registrar and two consultants independently graded the severity of each baby's illness without knowledge of the Baby Check score. The registrars assessed the babies at presentation while the consultants reviewed the notes. The consultants and registrars agreed about the need for hospital admission only about 75% of the time. The score's sensitivity and predictive values were similar to those of the registrars' grading. The score's specificity was 87%. Babies with serious diagnoses scored high, while minor illnesses scored low. The predictive value for requiring hospital admission increased with the score, rising to 100% for scores of 20 or more. The appropriate use of Baby Check should improve the detection of serious illness. It could also reduce the number of babies admitted with minor illness, without putting them at increased risk.

Baby Check is a score card developed to help parents and health professionals assess the severity of acute illness in babies under 6 months old.<sup>1 2</sup> It consists of 19 checks (seven questions about symptoms in the previous 24 hours and 12 simple examination signs), each carrying a score. The scores for positive checks are added together. The higher the total score, the sicker the baby.

Two versions have been produced: an illustrated booklet for parents and a card for doctors, health visitors, and midwives.<sup>1 3</sup> The card includes definitions, a table showing the chance of the baby being well or mildly ill, moderately ill, or seriously ill at different scores, and notes about low scoring conditions which require attention.

For parents, the total score is divided into four groups, a score between 0 and 7 means the baby is well or only mildly ill and is unlikely to need medical attention at present, a score between 8 and 12 that the baby is unwell but is not seriously ill, and that advice should be

sought from a doctor, health visitor, or midwife, a score between 13 and 19 that the baby is ill and needs to be seen by a doctor, and a score of 20 or more that there is a high chance of serious illness, and the baby needs to be seen by a doctor straight away. These cut offs were identified using Receiver Operating Characteristic curves and predictive values.

The score card simulates the clinical judgment of the original assessor, an experienced paediatrician, accurately. However, the predictive values, sensitivity, and specificity of the score apply only to the original study population. Before Baby Check can be adopted, it must be shown to be accurate in identifying seriously ill babies in other populations. This paper reports a field trial in which Baby Check was used to score babies presenting to hospital with an acute illness.

## Methods

During 44 weeks, 13 paediatric house officers at Addenbrooke's Hospital were asked to score every baby under 26 weeks old presenting for assessment of an acute illness. They received no instruction in the use of the score card. As soon after presentation as possible, without knowledge of the score, the duty paediatric registrar graded each baby's illness on a seven point scale, ranging from: 'Baby needs urgent hospital treatment for a life threatening condition' to: 'Well baby not requiring any special care or treatment'. The registrars' grading reflected the baby's state at the time of presentation. Two consultant paediatricians reviewed each baby's notes after discharge, using the same scale and without knowledge of the score. Their gradings took into account the investigation results, diagnosis, treatment, and outcome. For the analyses, these gradings were simplified into four categories, shown in table 1.

Baby Check's performance in identifying the babies the consultants considered needed admitting for observation or treatment was compared with that of the registrars' grading. Differences in sensitivity, specificity, and predictive accuracy between the score and the registrar's grading were explored using  $\chi^2$  analyses.

Table 1 Registrars' and consultants' gradings of illness severity. Results are number (%) of gradings

Grade of illness severity	Registrars (n=196)	Consultant A (n=259)	Consultant B (n=260)
(1) Has a serious illness needing hospital treatment	45 (23)	113 (44)	74 (28)
(2) Requires hospital admission for observation due to uncertainty about the severity of the illness	74 (38)	65 (25)	86 (33)
(3) Needs careful observation and treatment. Could be managed at home by a capable mother	47 (24)	59 (23)	65 (25)
(4) Mildly ill or well. Could be managed at home by any mother	30 (15)	22 (8)	35 (14)

University of Cambridge,  
Department of  
Paediatrics

A J Thornton  
C J Morley  
S J Green  
K A Walker  
J M Rennie

MRC Dunn  
Nutrition Unit,  
Cambridge  
T J Cole

Correspondence to:  
Dr C J Morley,  
University of Cambridge,  
Department of Paediatrics,  
Level E8,  
Addenbrooke's Hospital,  
Hills Road,  
Cambridge CB2 2QQ.

Accepted 10 August 1990

## Results

During the study 303/357 (85%) babies presenting to casualty were seen by the paediatric team and were thus eligible. The house officers scored 243 (80%). Nineteen babies presenting to the wards were also scored. Of these 262, 196 (75%) had their illnesses graded by the registrars: 172 (88%) were seen shortly after presentation, and in the remainder the registrar graded the illness from the notes a few hours later, using only the information available at the time of presentation. Fourteen babies who were sent home and 52 who were admitted did not have their illness graded by the registrar. Two hundred and fifty nine (99%) babies had their illness graded by consultant A and 260 (99%) by consultant B.

## SEVERITY OF THE BABIES' ILLNESSES

Of the babies scored, 227 (87%) were admitted (seven to intensive care) and 34 (13%) sent home (one not recorded). The median stay was two days, ranging from a few hours to 99 days; none died.

The babies had a broad range of diagnoses, from minor complaints such as nappy rash to serious illnesses such as meningitis.

Table 1 shows the registrars' and consultants' gradings of illness severity.

## PROFILE OF SCORES

The scores ranged from 0 to 57 (fig 1), with a median of 12 (10th and 90th centiles 0, 34). The median score for babies sent home was 3 (0, 7),

Figure 1 Profile of scores (n=262). The numbers of babies with each score are shown at the top of each column.

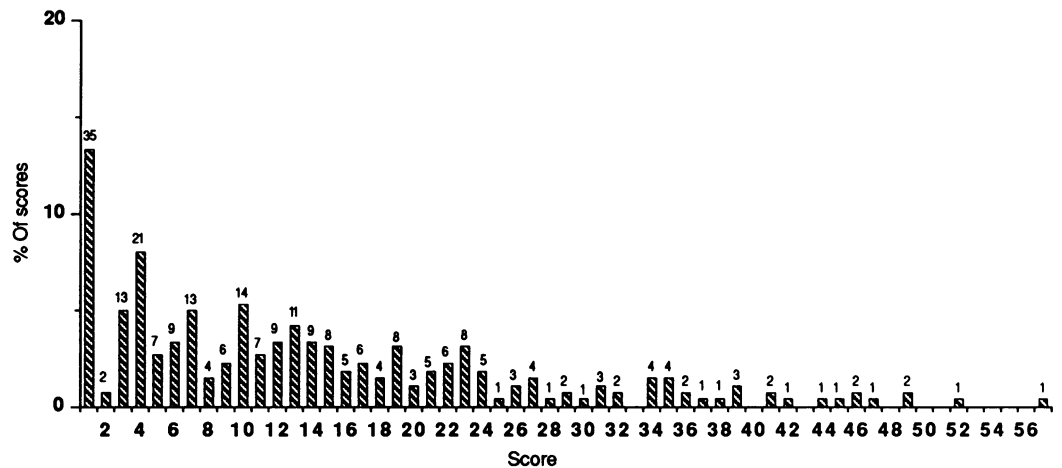


Table 2 Diagnoses and gradings of illness severity in a random selection of babies in each score group

Diagnosis	Score	Illness gradings		
		Registrars	Consultant A	Consultant B
<i>Scores of 0 to 7</i>				
Nasal obstruction	0	3	3	3
Vomiting—no cause found	0	4	3	3
Pyloric stenosis	0	3	1	1
Minor head injury	0	4	4	3
Asymptomatic heart murmur	0	4	4	4
Vomiting—no cause found	4	3	3	4
Well, crying baby	4	—	4	4
Drowsy—no cause found	5	2	2	2
Upper respiratory tract infection	7	3	3	3
<i>Scores of 8 to 12</i>				
Acute otitis media	8	—	2	2
Upper respiratory tract infection	9	4	3	3
Bronchiolitis	10	3	1	3
Viral infection	10	3	1	2
Pyloric stenosis	10	1	1	2
Anal fissure	11	—	2	3
Staphylococcal skin infection	12	1	1	1
Bronchiolitis	12	2	2	2
<i>Scores of 13 to 19</i>				
Viral infection	13	3	3	2
Gastroenteritis	14	3	3	4
Bronchiolitis	14	—	1	3
Septicaemia	15	2	1	1
Bronchiolitis	15	3	1	2
Upper respiratory tract infection	15	3	4	3
Viral infection	16	2	3	2
Convulsion	18	—	1	2
Bronchiolitis	19	2	1	3
Renal dysplasia	19	2	1	1
<i>Scores of 20 or more</i>				
Bronchiolitis	24	2	2	1
Bronchiolitis	27	3	2	2
Bronchiolitis	27	1	1	1
Urinary tract infection	29	2	1	1
Viral pneumonia	31	2	1	1
Vomiting—no cause found	34	—	3	2
Intussusception	34	1	1	1
Near miss cot death	37	1	1	1
Bacterial pneumonia	52	1	1	1
Irreducible inguinal hernia	49	1	1	1

and for those admitted to paediatric wards and to intensive care, 13 (3, 34) and 30 respectively. Those graded by the registrars as needing hospital treatment (grade 1) had a median score of 27 (10, 46). Those graded as well or mildly ill (grade 4) had a median score of 4 (0, 16).

*The score groups*

Of the 262 scores, 100 (38%) were between 0 and 7, 40 (15%) 8 to 12, 51 (20%) 13 to 19, and 71 (27%) 20 or more.

Table 2 shows the diagnoses and gradings of illness severity for a random selection of babies in each score group.

SCORES FOR DIFFERENT DIAGNOSES

Table 3 shows the distribution of scores for some well defined diagnoses. The highest scores (52, 57) were for babies with bacterial pneumonia.

THE SCORE COMPARED WITH THE GRADINGS OF ILLNESS SEVERITY

A registrar and both consultants graded the illness in 193 (74%) cases. Table 4 shows comparisons of the gradings. Table 5 shows the score groups compared with the gradings. Data for 193 babies are presented to facilitate comparisons using the same cohort throughout. Using all the available data gave very similar results.

*Concordance between the paediatricians*

The three paediatricians' gradings were compared for concordance using the percentage

Table 3 Scores for different diagnoses

Diagnosis	No of babies	Median	Range
Bacterial pneumonia	2	54	52-57
Meningitis	3	35	22-42
Intussusception	4	31	30-45
Urinary tract infection	3	29	11-36
Inguinal hernia	8	23	13-49
Bronchiolitis	48	22	7-49
Pyloric stenosis	9	13	0-39
Gastroenteritis	20	13	2-26
Convulsion	5	12	0-18
Abscess/cyst	8	10	0-44
Upper respiratory tract infection	36	7	0-31
Apnoea	14	5	0-46
Constipation	4	4	3-7
Well/teething	14	4	0-15

Table 4 Comparison of the registrars' and consultants' gradings (1-4) of illness severity (n=193)

	Consultant A illness grading			
	4	3	2	1
Registrars' illness grading	4	8	12	3
	3	2	18	12
	2	2	12	23
	1	—	1	6
	37			
	Consultant B illness grading			
	4	3	2	1
Registrars' illness grading	4	9	16	4
	3	6	22	15
	2	2	11	31
	1	—	1	13
	30			30
	Consultant A illness grading			
	4	3	2	1
Consultant B illness grading	4	4	10	1
	3	7	19	15
	2	1	12	22
	1	—	2	6
	28			55

Table 5 Score groups compared with registrars' and consultants' gradings of illness severity (n=193)

Score group:	Registrars' illness grading			
	4	3	2	1
0 to 7	23	23	19	3
8 to 12	3	5	16	5
13 to 19	3	15	16	6
20 or more	—	3	23	30
Score group:	Consultant A illness grading			
	4	3	2	1
0 to 7	11	22	17	18
8 to 12	—	8	7	14
13 to 19	1	10	10	19
20 or more	—	3	10	43
Score group:	Consultant B illness grading			
	4	3	2	1
0 to 7	14	30	20	4
8 to 12	2	8	11	8
13 to 19	1	11	15	13
20 or more	—	1	17	38

agreement and the  $\kappa$ ,<sup>4 5</sup> with a cut off between grades 2 and 3 (that is, whether or not the baby needed admitting to hospital). The results were: registrars compared with consultant A, 74% agreement,  $\kappa=0.42$ ; registrars compared with consultant B, 81% agreement,  $\kappa=0.60$ ; and consultant A compared with consultant B, 78% agreement,  $\kappa=0.50$ .

*Concordance between the score and the paediatricians*

The same procedure was used to compare the score and the illness gradings, using the same cut off between grades 2 and 3, and a score of 13 or more. The results were: score compared with registrars, 67%,  $\kappa=0.34$ ; score compared with consultant A, 64%,  $\kappa=0.28$ ; and score compared with consultant B, 71%,  $\kappa=0.42$ .

Thus the agreement between the score and the paediatricians was slightly less good than among the paediatricians.

SENSITIVITY

*Babies requiring hospital treatment (grade 1)*

There was no significant difference between the sensitivity of the score and that of the registrar in identifying babies the consultants graded as needing hospital treatment. The sensitivity of the registrars was 37/94 (39%) for consultant A and 30/63 (48%) for consultant B. The sensitivity of a score of 20 or more was 43/94 (46%) and 38/63 (60%) respectively.

*Babies requiring hospital admission (grades 1 and 2)*

A score of 13 or more identified significantly fewer of the babies graded by the consultants as requiring admission than the registrar did ( $p<0.05$ ). The sensitivity of the registrars was 103/138 (75%) for consultant A and 104/126 (83%) for consultant B. The sensitivity of a score of 13 or more was 82/138 (59%) and 83/126 (66%) respectively.

SPECIFICITY

The specificity of the registrars identifying babies the consultants graded as mildly ill (grade 4) was 8/12 (67%) and 9/17 (53%) respec-

tively. The specificity of a score of 0 to 7 was 11/12 (92%) for consultant A and 14/17 (82%) for consultant B.

A low score thus had a higher specificity than the registrars' grading. The difference was not significant due to the small numbers.

**PREDICTIVE VALUES**

*Individual scores*

Figure 2 shows the predictive values of each score for each grade of illness, averaged over the two consultants (n=259). The higher the score the sicker the baby. The chance of needing hospital admission or treatment increased with the score. For example, the predictive value for needing hospital treatment was low at low scores, rising to 67% at a score of 20 and 100% at a score of 28.

Few babies scoring over 8 were graded as mildly ill. The number graded as needing careful observation at home (grade 3) also decreased as the score increased.

*Score groups*

**Babies requiring hospital treatment (grade 1): scores of 20 or more**—A score of 20 or more was as predictive as the registrars' grading in identifying babies needing hospital treatment. The registrars' grading had predictive values of 37/44 (84%) for consultant A and 30/44 (68%) for consultant B (table 4). The predictive values of a score of 20 or more were 43/56 (77%) and 38/56 (68%) respectively (table 5).

**Babies requiring hospital admission (grades 1 and 2): scores of 13 or more**—The predictive values of the registrars' grading were 103/118 (87%) for consultant A and 104/118 (88%) for consultant B. The predictive value of a score of 13 or more

for being graded as needing admission was 82/96 (85%) for consultant A and 83/96 (86%) for consultant B. These results were not significantly different.

**Babies graded as well or mildly ill (grade 4): scores of 0 to 7**—The registrars' grading a baby as being mildly ill or well had a predictive value for being graded the same of 8/29 (28%) for consultant A and 9/29 (31%) for consultant B (table 4). The predictive values of a score between 0 and 7 were 11/68 (16%) and 14/68 (21%) respectively (table 5). These results were not significantly different.

**FALSE NEGATIVES AND FALSE POSITIVES**

These were identified using the registrars' grading, because this was recorded at the same stage of the illness as the score. Misclassifications arising from the consultants' gradings were very similar.

*False negatives: babies with low scores requiring hospital treatment (grade 1)*

Three babies (7%) graded 1 had scores between 0 and 7. Their diagnoses were apnoeic episode, hydrocephalus, and pyloric stenosis (only scored for vomiting). Five (11%) scored 8 to 12. Their diagnoses were thyroglossal cyst, proctitis, viral infection (initially thought to be meningitis), pyloric stenosis, and staphylococcal skin and eye infection.

*False negatives: babies with low scores requiring hospital observation (grade 2)*

Nineteen babies (16%) graded 2 scored between 0 and 7. The diagnoses made were: apnoeic episodes (n=5), upper respiratory tract infection

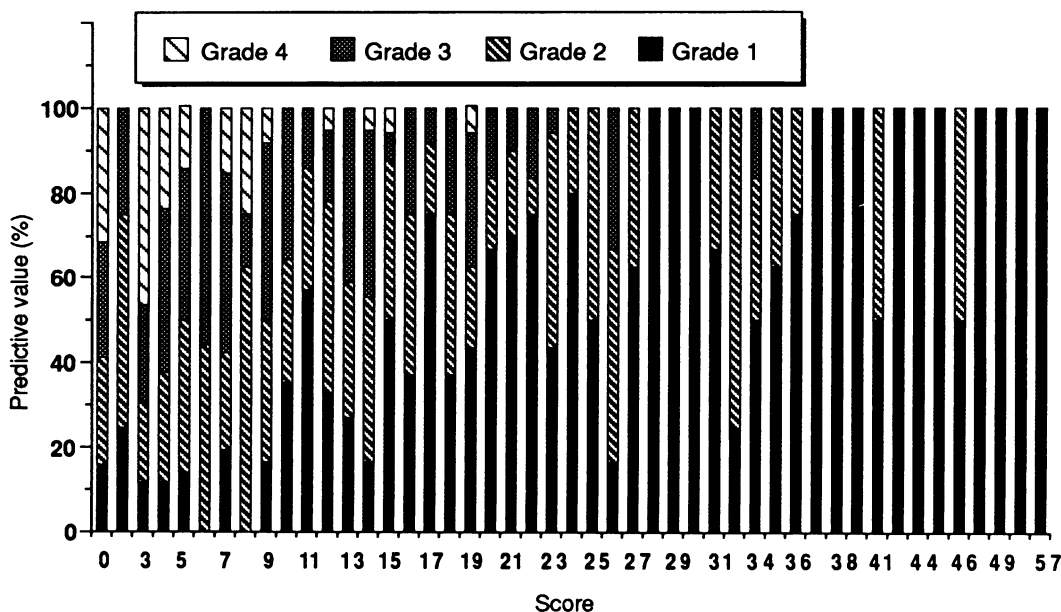


Figure 2 Predictive value of each score for each grade of illness (n=259). Three babies did not have their illness graded by both paediatricians; they scored 16, 34 and 39. The four grades of illness are defined as: Grade 1, has a serious illness needing hospital treatment. Grade 2, requires hospital admission for observation due to uncertainty about the severity of the illness. Grade 3, needs careful observation and treatment. Could be managed at home by a capable mother. Grade 4, mildly ill or well. Could be managed at home by any mother.

(n=3), bronchiolitis (n=2), gastro-oesophageal reflux (n=2), and one case each of convulsion, non-accidental injury, infected umbilicus, vomiting (no cause found), gastroenteritis, drowsiness (no cause found), and viral infection.

*False positives: babies with high scores graded as well or mildly ill (grade 4)*

Three babies (10%) graded 4 had scores of 13 or more. None scored over 19. The diagnoses made were: proctitis, upper respiratory tract infection, and jaundice.

### Discussion

These results show that Baby Check provides an accurate means of grading the severity of acute systemic illness in babies. The predictive value for serious illness increases with the score. Babies with serious diagnoses score high and those with minor illnesses score low. The score has a high specificity. The sensitivity and predictive value are similar to a paediatric registrar's grading of illness severity.

There is no gold standard to measure illness severity, and the field trial shows how difficult it is, even for experienced paediatricians. They only agreed about whether a baby needed admitting three quarters of the time. Many of the babies had non-specific illnesses which were difficult to assess, and about two thirds of admissions were because of uncertainty about the severity of the illness. The agreement between the score and the paediatricians was only slightly less good than that between the paediatricians themselves.

In a study of babies presenting to hospital it might seem strange that over half scored less than 13. But many of the babies were not seriously ill. About 40% were graded as suitable for home management, provided the mother could cope. The number admitted reflects hospital policy—babies presenting for acute assessment are admitted for observation unless they are obviously well.

Although the use of 'cut offs' is necessary to explore the score's sensitivity and specificity, they are a crude test of its accuracy. The score provides a continuous measure of illness severity. The risk of serious illness increases with the score (fig 2). The irregularities in the predictive values are due partly to the small numbers of babies at the higher scores. The predictive values from the original population shown on the professionals' version of Baby Check are smoothed.<sup>1</sup>

The common diagnoses gave rise to a range of scores and paediatricians' gradings, reflecting the variation in the severity of disease. Baby Check is thus useful in grading illness severity in variable conditions such as bronchiolitis. Serious conditions such as meningitis and pneumonia always scored 20 or more. Diagnoses such as jaundice and convulsions were associated with lower scores, reflecting the fact that they often have few systemic signs.

Scores of 20 or more had a predictive value for being graded as needing admission of over

95%, with over 70% graded as needing hospital treatment. Scores in this range were as effective as the registrars in identifying the sickest babies—those graded as requiring treatment for a known serious condition. Most babies with scores over 19 had serious diagnoses.

A score of 13 or more was slightly less sensitive than the registrars in identifying babies graded as needing admission (grades 1 and 2), identifying approximately two thirds, with a positive predictive value of around 85%.

Baby Check must not miss seriously ill babies. The low scoring babies graded as needing admission fell into two groups: some had specific conditions such as convulsions, abscesses, and injuries that have few systemic signs but still need investigation. The score card is not designed to assess such conditions, which are easily recognisable. It includes a warning that they may give a low score but still warrant careful assessment. The second group comprised those admitted because of uncertainty about how the illness might progress. These babies had a wide range of illnesses and scores. These two factors also accounted for the poor negative predictive value of the lower scores.

One fifth of the babies graded by the registrars as needing hospital treatment (grade 1) scored less than 13. According to the consultants, however, the registrars overestimated the number needing treatment in a quarter of babies.

The specificity of the score was higher than that of the registrars. Some 87% of the well or mildly ill babies scored less than 8.

The false positive rate was low. Only three babies graded by the registrars as mildly ill scored over 12. None scored over 19. All were graded more seriously by the consultants.

Predictive values and sensitivities apply only to the population from which they were derived. In this trial Baby Check was tested by multiple observers in an environment that differed from the original one. The predictive values from the original population, shown on Baby Check,<sup>1</sup> and the values observed in this study differ mainly at the lower scores. The original study included a large community cohort,<sup>1 2</sup> so that most of the low scoring babies were well.

Baby Check is designed for use by parents at home, general practitioners, health visitors, and midwives in the community and junior doctors in hospital. In each of these environments the prevalence of illness is different, which will affect the interpretation of the lower scores. At home most babies are well and achieve low scores.<sup>3</sup> A baby with a low score at home has more than a 90% chance of being well or mildly ill.<sup>1 2</sup> In hospital many low scoring babies have conditions (such as a convulsion) which obviously need attention, as demonstrated by the poor predictive value of the low scores in this study.

Although many of the babies scoring 13 to 19 were graded as requiring admission once they had presented to hospital, few were thought to need treatment for a known serious illness, and if seen by a general practitioner some of these babies might have been managed at home with careful review. Scores of 20 or more are rare in

the community,<sup>1-3 5</sup> however, and this study confirms that babies who achieve them warrant urgent assessment wherever they are seen.

The score provides an objective means of grading a baby's illness, which can be used in association with clinical judgment to decide what action should be taken. It should help doctors assess the severity of acute systemic illness in babies. The score is of particular value in identifying the well and the most seriously ill babies, and should assist in the management of the large group of babies who are currently admitted because of uncertainty about the severity of the illness. Provided the home circumstances and previous history were satisfactory, some of those with low scores could be sent home and kept under review by the mother and the general practitioner using Baby Check,<sup>3 6</sup> returning to hospital if the score increased.

The appropriate use of Baby Check in hospital and in the community should improve the detection of serious illness. It could also reduce

the number of babies admitted with mild illness, without putting them at increased risk.

We thank the Baby Illness Research Project Appeal of the Foundation for the Study of Infant Deaths for financial support, the mothers and babies of Cambridge, the paediatric senior house officers, registrars, and consultants at Addenbrooke's Hospital for their help, and Dr PH Hewson and MA Fowler for their contribution to the original study.

Copies of the Baby Check score card and booklet can be obtained from the authors.

- 1 Morley CJ, Thornton AJ, Cole TJ, Hewson PH, Fowler MA. Baby Check: a scoring system to grade the severity of acute systemic illness in babies under 6 months old. *Arch Dis Child* 1991;**66**:100-5.
- 2 Cole TJ, Morley CJ, Thornton AJ, Fowler MA, Hewson PH. A scoring system to quantify illness in babies under six months of age. *Journal of the Royal Statistical Society A* 1990 (in press).
- 3 Thornton AJ, Morley CJ, Green SJ, Cole TJ, Walker KA, Bonnett JM. Field trials of the Baby Check score card: mothers scoring their babies at home. *Arch Dis Child* 1991;**66**:106-10.
- 4 Kramer MS, Feinstein AR. Clinical biostatistics LIV. The biostatistics of concordance. *Clin Pharmacol Ther* 1981;**29**:111-23.
- 5 Landis JR, Koch GG. The measurement of observer agreement for categorical data. *Biometrics* 1977;**33**:159-74.
- 6 Morley CJ, Green SJ, Thornton AJ, Cole TJ. Field trials of the Baby Check score card in general practice. *Arch Dis Child* 1991;**66**:111-4.