

Evaluation of a near-patient test for C-reactive protein used in daily routine in primary healthcare by use of difference plots

BJARNE STEEN DAHLER-ERIKSEN,^{1,2*} JENS FLENSTED LASSEN,¹ PER HYLTOFT PETERSEN,³
ERIK DALSGAARD LUND,¹ TORSTEN LAURITZEN,² and IVAN BRANDSLUND¹

We have assessed the technical performance and robustness of NycoCard[®] CRP Whole Blood, a near-patient test for C-reactive protein (CRP), when used in realistic daily routine situations in general practice clinics (GPC). Thirteen GPCs participated, five of them with technician staff. From 898 patients, split-sample measurements for CRP were made. Results from GPCs were compared with results from a turbidimetric laboratory method, traceable to international reference preparations (IFCC CRM 470). Results were evaluated in difference plots where the expected distribution, due to an estimated analytical variation, was compared with measured differences. Of all difference points, 91.5% (n = 819) were within a 95% prediction interval based on the imprecision of both methods. Mean bias (95% confidence interval) was -0.3 mg/L (-0.9 to 0.3). No differences in analytic quality were found between GPCs with technician staffs and GPCs without, and between test results obtained within the first and second week, compared with the rest of the study period. We find the test as good when used in GPCs as could be expected from laboratory testing, and consequently robust, which is a necessity for use in routine situations in general practice. General application of difference plots in test evaluations are discussed in detail.

The availability of near-patient tests has expanded rapidly during the past decade because of the technical evolution and the increased pressure for rapid diagnosis and treatment [1]. In the US, ~20% of all testing are near-patient

tests [2]. For doctors and healthcare planners, additional use of near-patient testing in general practice is desirable for taking care of patients discharged earlier from the hospital and for the surveillance of patients being moved from outpatient clinics at hospitals to general practice.

The efficacy and safety of a new near-patient test should ideally be demonstrated before the introduction of the test in routine situations in general practice. A guideline for implementation of near-patient tests emphasized that [3]: "It is likely that the equipment will be operated by staff who are not trained as analysts and particular importance should be attached to the robustness of the analytical system: i.e., not what its performance is in the best conditions of operation, but rather what will be achieved in the conditions in which it will actually be used." The term robustness is defined as the ability to yield acceptable results of measurements in spite of deviation from details of the measurement procedure [4].

In general practice, patients with infectious diseases constitute a major part of all consultations [5]. General practitioners (GPs) use laboratory tests when assessing these patients.⁴ One such test is for C-reactive protein (CRP), a marker of the acute-phase response [6–8]. Because the plasma concentration of CRP increases rapidly after stimulation and decreases rapidly with a short half-life, CRP can be a very useful tool in diagnosing and monitoring infections and inflammatory diseases [6–8].

In Denmark, all CRP measurements from primary and secondary healthcare are performed at hospital laboratories. In the service area of Vejle County Central Hospital (Vejle Hospital), GPs use a CRP measurement in one of 20 consultations and request a CRP measurement for one in three blood samples mailed to the laboratory [9].

¹ Department of Clinical Chemistry, Vejle County Central Hospital, DK-7100 Vejle, Denmark.

² Department of General Practice, University of Aarhus, Aarhus, Denmark.

³ Department of Clinical Chemistry, Odense University Hospital, Denmark.

*Author for correspondence. Fax + 45 75 82 18 14.

Received November 25, 1996; revision accepted May 9, 1997.

⁴ Nonstandard abbreviations: GP, general practitioner; CRP, C-reactive protein; GPC, general practice clinic; and CI, confidence interval.

Now simple and rapid methods for CRP measurement have been developed, suitable as near-patient tests in general practice. Having access to a rapid CRP result while the patient is still in the doctor's office could be valuable, avoiding unnecessary prescription of an antibiotic, by helping the doctor distinguish between viral and bacterial infection [1, 10, 11]. The technical performance of some of these tests is well evaluated when used in a laboratory by experienced technicians [12, 13] but only few studies have been carried out with the test in routine situations in general practice [14, 15]. These studies have compared the CRP results with the CRP result from a laboratory nearby as the "true" CRP value. But external quality assessments for central laboratories have shown a major interlaboratory variability of CRP measurements [16]. Reference laboratories participating in studies evaluating the technical performance of near-patient tests must be able to document their methods being traceable to BCR/CAP/IFCC CRM 470 (international reference preparation) to ensure the analytical quality for the "true" CRP value. Moreover, the robustness of the analytical system in terms of "importance of the analytical experience for the person operating the test" or "time needed for personnel before proper test performance" must be evaluated before implementation [3]. None of the previous studies evaluating near-patient tests for CRP measurements, when used in general practice, fulfills these requirements [14, 15].

The aim of this study was to evaluate the technical performance, especially to test the robustness, of a near-patient test for CRP measurements when used in daily routine in general practice, with measurements carried out by technician staff or unskilled laboratory personnel, and compare results with a documented high-quality laboratory method. We aimed to make a clear presentation of the results by use of difference plots, with a 95% prediction interval based on estimated and assumed obtainable analytical imprecision.

Materials and Methods

GENERAL PRACTICE CLINICS (GPCs)

The catchment area of Vejle Hospital consists of 120 000 persons and contains 41 GPCs with 84 GPs, all serviced at the laboratory at Vejle Hospital. Thirteen GPCs (47 GPs) were, after randomized selection, asked to participate in the evaluation, and all 13 accepted. Five of the GPCs had technician staff. In the other eight GPCs, blood sample drawing, analytical procedures, and test reporting were made by unskilled laboratory personnel (nurses, secretary staff, or the GPs themselves). All the GPCs had received an introductory visit with demonstration from the manufacturer before the study period, but besides this visit, no training period was included.

PATIENTS

Patients were included during a 2-month period in February and March 1995. All patients for whom a blood test

for CRP was ordered by the GP for a clinical purpose were included in the study. Of the 909 patients included, 551 (60.6%) were female, with a median age (range) of 58.5 years (6–97), and 358 (39.4%) were male, with a median age of 59.3 years (10–98).

CRP MEASUREMENTS IN GPCS

All blood samples were drawn by venipuncture with the closed Monovette system (Sarstedt) with heparin as anticoagulant. With a capillary straw, 25 μ L of whole blood for near-patient testing was removed from the test tube and the tube was then routinely mailed to the laboratory at Vejle Hospital for CRP measurements.

NycoCard[®] CRP Whole Blood (Nycomed Pharma) was investigated. The test system is based on an immunometric principle and consists of (a) a liquid for sample dilution and lysis of blood cells, (b) a test card with six holes containing CRP-specific monoclonal antibodies coated to a membrane, (c) a conjugate solution with monoclonal CRP antibodies coupled to small gold particles, and (d) a washing solution.

The CRP measurement was performed by diluting the 25 μ L of whole blood in the capillary straw in 1000 μ L of dilution liquid. After mixing and 45 s of lysis time, 25 μ L of this diluted sample was applied to a test hole on the test card. After allowing the liquid to soak into the membrane, one drop of the conjugate solution was applied and afterwards one drop of the washing solution was applied to remove conjugate solution in surplus. The gold particles coupled to the CRP antibodies in the conjugate give the membrane a purple-reddish color, and the CRP value is measured from the intensity of the color at the membrane. The color is either measured quantitatively with a color densitometer (NycoCard Reader) or semiquantitatively by visual comparison with a reference color chart corresponding to CRP values of 10, 25, 50, 100, and 200 mg/L. In this study, visual reading of the test results was evaluated.

The person performing the test in GPCs was asked to state the results as "best guess" within 5 mg/L from 10–50 mg/L, within 10 mg/L from 50–100 mg/L, and within 25 mg/L for values >100 mg/L. Possible results were <10, 10, 15, 20, 25, 30, 35, 40, 45, 50, 60, 70, 80, 90, 100, 125, 150, 175, 200, 225, and >225 mg/L.

For estimation of CVs of the test kit, three of the GPCs were asked to make the CRP measurements as duplicate measurements. The GPCs were asked to arrange the testing with two different employees, ensuring the independency of the double measurements. The CV for 117 duplicated measurements was 0.7% (SD = 0.14, mean CRP value = 21.12 mg/L). In a previously published laboratory evaluation of this test, the CV was from 10% to 15% between series for values at 25 mg/L (n = 27) and 130 mg/L (n = 27), when all measurements were performed by 11 experienced technicians and the reading of the color response was measured with a color densitometer [13]. Thus, the CV found in the three GPCs is

surprisingly low, and it is likely that the readings were not made independently, but rather were a result of a consensus view between two employees. For the further calculations in this study, we have used an approximation for the CV at 15%, as obtainable when the test kit is used in general practice, based on the laboratory evaluation [13]. It seems that this is the best that can be expected in general practice.

The test allows the measurements to be carried out in whole blood, but presents the results as serum values assuming a hematocrit of 0.40. CRP is then slightly overestimated when hematocrit is low and slightly underestimated when hematocrit is high [13]. The manufacturer recommends to correct for hematocrit when value is >0.55 . However, when an infectious patient is seen by a GP, values of hematocrit are only seldom available, and if the GP has to perform a hematocrit every time he orders a bedside CRP, the idea of using bedside tests will disappear. So in daily routine use we do not expect the GP to correct for hematocrit, and as we wanted to evaluate the test kit while used in daily routine in general practice, we have made the same decision.

CRP MEASUREMENTS AT THE LABORATORY

CRP was determined by turbidimetry with a Hitachi 717 analyzer with antibodies (cat. no. 67128) and buffers (cat. no. 67179) from Orion Diagnostica. The calibrator was from Dako (code no. X 0923). The accuracy of the calibration was checked with the international reference preparation for immunochemical measurements, BCR/CAP/IFCC CRM 470 (Commission of the European Communities, Bruxelles, Belgium) [17]. The reference preparation has an assigned value of 39.2 mg/L [confidence interval (CI) 95%: 37.3–41.2], and when checking the calibration four different times in the study period, we found 38 mg/L. As control samples we used a pool of patient serum with a low CRP value and a pool of patient serum with a high CRP value. In the study period the following means and coefficient of total analytical variation (CV) for a specified number of control measurements were found: low pool: mean = 29.7 mg/L and CV = 5% ($n = 185$), and high pool: mean = 136.1 mg/L and CV = 5% ($n = 51$). Results were given quantitatively except for CRP values <10 mg/L.

STATISTICAL ANALYSIS

CRP results from general practice and from the laboratory were compared by using four different methodologies:

1) Results as “best guess” from general practice were compared with the quantitative results from the laboratory divided in intervals matching the answer from the general practice.

2) The mean of difference, defined as the mean difference between CRP measurements performed with the near-patient test in general practice and CRP measurements performed at the laboratory (near-patient test result – laboratory results; unit: mg/L), and the standard

deviation for the mean of differences, were calculated. A 95% CI for the mean of differences was calculated by using the t -distribution, and a 95% CI for the SD of differences was calculated by using the χ^2 -distribution. These values were calculated for all participating GPCs, for five intervals of laboratory CRP values (<10 , 10–25, 26–50, 51–100, and >100 mg/L), for GPCs with or without technician staff, for GPCs with more or less than 50 tests performed in the study period, and for test results performed within the first week or two of the study period compared with results from the rest of the period. CRP values read as <10 mg/L in the GPCs and measured <10 mg/L at the laboratory were defined as 9 mg/L in the calculations.

3) Linear regression was used to compare the paired results from our study with results from other studies.

4) The paired results were evaluated with difference plots. A 95% prediction interval is calculated, expressing the interval within which we would expect 95% of the data points to be found, on the basis of knowledge about the analytical variation (CV) of the two tests ($CV^2_{\text{difference}} = CV^2_{\text{lab}} + CV^2_{\text{test}}$). Knowing the CVs for the two tests, it is possible before actually performing the practical part of the test evaluation to set up a prediction interval for which a certain fraction of the difference points are expected to be distributed. Such a 95% prediction interval is in contrast to the 95% limits of agreement, described by Bland and Altman, which is based on calculations of the actually measured difference between the two methods and thereby describes the 95% interval for measured differences [18, 19]. The calculation of the 95% prediction interval is based upon the analytical variation for the laboratory method, $CV_{\text{lab}} = 5\%$, and the CV for the near-patient test, CV_{test} composed by the analytical variation $CV_{\text{test-analytic}} = 15\%$, approximated from the laboratory study [13], and a variation for the readings, $CV_{\text{test-read}}$. When using a semiquantitative test where test results are reported in intervals, we normally have no specific knowledge about the possible values of the test within the interval, but we can assume the value to lie anywhere within the interval, which gives us a rectangular distribution of the test results [20]. The CRP results from general practice was given within a “best guess” interval (within 5 mg/L in the interval from 10 to 50 mg/L, within 10 mg/L from 50 to 100 mg/L, and within 25 mg/L for values >100 mg/L). As example: If a “true value” of CRP is within the interval of 17.5–22.5 mg/L, the answer 20 mg/L is considered correct for all CRP values within this interval, but as we do not know the specific value within the interval, a reading variation is then introduced. Assuming the rectangular distribution for this reading variation, a standard deviation can be transformed from the formula: $a/(2 \times 3^{1/2})$ where a is the length of the whole interval (here 5 mg/L) [20]. (Further considerations and calculations of the 95% prediction interval are described in detail in the *Appendix*.) The between-method differences are plotted both with the result for the laboratory

method as abscissa (Fig. 1A) and with the average results for the two methods as abscissa (Fig. 1B), the latter as suggested by Bland and Altman [18, 19]. Further difference plots are presented logarithm-transformed (ln-transformed) (Fig. 2).

ETHICS

The procedures followed in this study were in accordance with the second Declaration of Helsinki (amended in 1989), the Danish law on biomedical research, and the Scientific Ethical Committee system of October 1, 1992. According to this, for research on existing data without intervention and with the purpose of technical and medical quality assurance, an approval by a scientific ethical committee is not needed. The local scientific ethical committee was informed about the study. All included patients had a CRP measurement ordered routinely by their GP, and blood from this routine test was used to evaluate the quality of the near-patient tests.

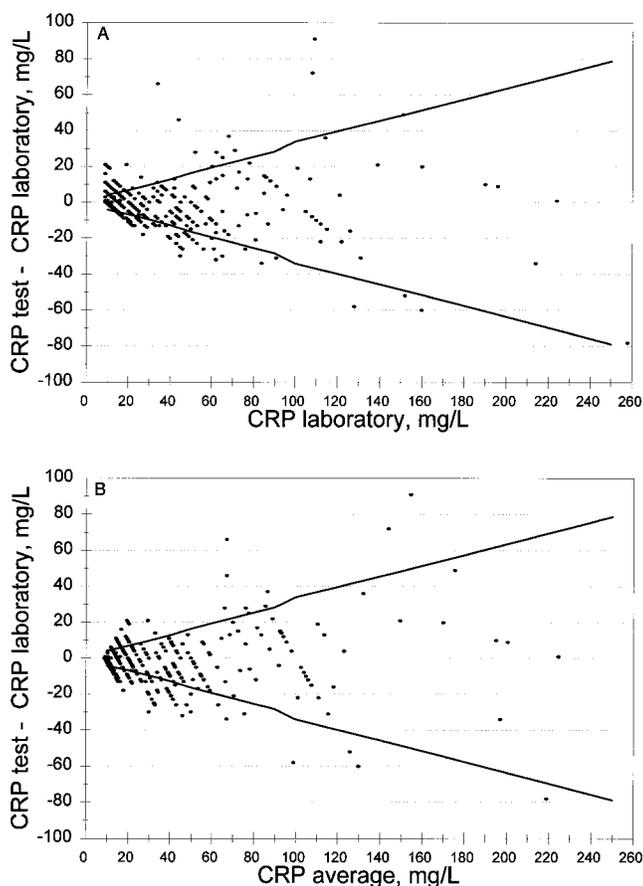


Fig. 1. Difference plots comparing results of CRP measured with a near-patient test (test) in general practice with results measured by turbidimetry in the laboratory.

Lines indicate the 95% prediction interval, an interval where 95% of the data points are expected to be found, having knowledge about the CVs of the two methods. Differences are plotted against (A) laboratory measurement and (B) average results of the two measurements.

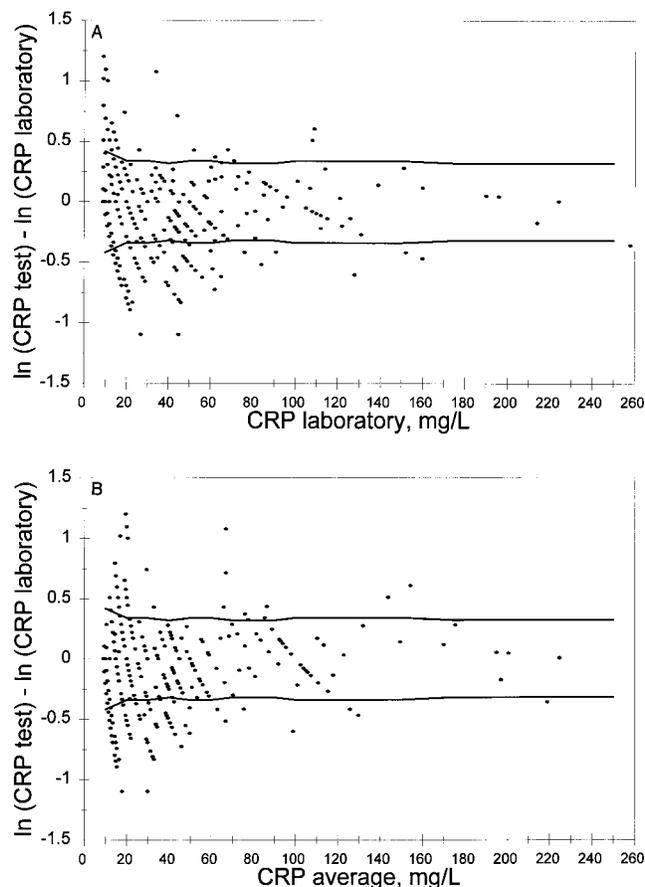


Fig. 2. Logarithm-transformed difference plot of the same data as illustrated in Fig. 1.

Results

In total, 909 CRP measurements were performed in the 13 GPCs during the study period. In 11 cases it was not possible to compare test results: For eight of these samples a CRP measurement was not ordered to the laboratory by the GP, one sample was too old for measurement at the laboratory because of incorrect mailing procedure, one sample was never received at the laboratory, and one sample failed in the analytical procedure in the laboratory. The material then comprises 898 samples useful for comparison of test results.

COMPARED TEST RESULTS

Results from all test samples are shown in Table 1. Quantitative results from the laboratory are divided into intervals matching the "best guess" results for the test kit. In 61% ($n = 549$) the results of both methods were within the same "best guess" interval and in 82% ($n = 736$) the results were within the same interval or in an adjacent interval. Approximately 70% ($n = 634$) of all CRP values measured in the laboratory were within the normal reference interval (≤ 10 mg/L), 20% ($n = 176$) were > 20 mg/L, and $\sim 10\%$ ($n = 92$) were high (≥ 50 mg/L).

Results from all 13 GPCs are shown in Table 2. The

Table 1. Paired results from two methods of CRP measurement.

Near-patient test CRP measurements, mg/L	Turbidimetric CRP measurements, mg/L										
	<10	10	15	20	25	30	35	40	45	50	60
	10–12	13–17	18–22	23–27	28–32	33–37	38–42	43–47	48–55	56–65	
<10	490	17	20	5	1						
10	72	8	12	4	1						
15	24	8	12	5	3	2		1			
20	5	5	9	7	12	3	2	3	3		
25	1		7	5	6	2	3	1	2		
30	2	2		1	2	2	3	1		4	2
35					1	2	2	3	2	2	2
40			1	1	1	3	0	5	5	1	
45						3	4	1			2
50								1	1	5	3
60							1			3	3
70											1
80									1		3
90								1			2
100						1					
125											
150											
175											
200											
225											
Total	594	40	60	28	26	11	14	15	18	20	19
Cumulative %	66.1	70.6	77.3	80.4	83.3	84.5	86.1	87.8	89.8	92.0	94.1

CRP values from 898 samples measured in a general practice with a near-patient test compared with a turbidimetric laboratory method. Quantitative results from the laboratory are divided into intervals matching the results from the test kit.

overall mean of differences (95% CI) between the paired CRP measurements was -0.3 mg/L (-0.9 to 0.3 mg/L) and the SD of the differences was 9.6 mg/L. A tendency towards a higher SD of difference for GPCs with a higher mean value of CRP is seen [$r = 0.50$ between the SD of difference and the mean value of CRP found in the GPCs ($n = 13$)(Table 2)].

Using linear regression to evaluate the overall number of split-sample measurements ($n = 898$), we found $r = 0.94$.

In Table 2 the results from all measurements are shown, grouped in five intervals of CRP values (<10, 10–25, 26–50, 51–100, and >100 mg/L). For all five CRP intervals the mean of differences is within -4.5 and 0.6 mg/L. The SD of difference depends upon the numerical value of the CRP results, which means that it is not correct to use a single value of SD to describe the test.

EFFECT OF ANALYTICAL EXPERIENCE

Cumulated data from the five GPCs with technician staff and from the eight GPCs with unskilled laboratory personnel are shown in Table 2. No statistical difference was found between the two groups according to the mean of difference, but the SD of difference was, with statistical significance, slightly higher among the nontechnicians than among technicians.

EFFECT OF A LEARNING PERIOD

To assess the importance of a learning period before routine use of the test, results obtained within the first week or two were compared with test results obtained in the rest of the study period (Table 2). No statistical differences were demonstrated between the two groups comparing the mean of difference. When we compared the SD of difference, measurements performed in the first weeks of the study agreed, with statistical significance, a little more precisely than in the rest of the study period. When we compared results for GPCs with >50 CRP measurements with GPCs with <50 CRP measurements in the study period, no statistical differences were demonstrated on the mean of difference. When we compared the SD of the difference, measurements performed in GPCs with a low number of CRP measurements agreed, with statistical significance, slightly better (Table 2).

DIFFERENCE PLOT

The measured difference between the two methods (the between-method difference) is plotted against the results of the laboratory method in Fig. 1A and the average of the two measurements in Fig. 1B. In both figures, an upper and lower limit for the 95% prediction interval is shown. The 95% prediction interval becomes wider for higher CRP values, illustrating the effect of an inconstant stan-

Table 1. Continued.

Turbidimetric CRP measurements, mg/L										
70	80	90	100	125	150	175	200	225	Total	Cumulative %
66-75	76-85	86-95	96-112	113-137	138-162	163-187	188-212	213-		
									533	59.4
									97	70.2
									55	76.3
									49	81.7
									27	84.7
									19	86.9
									12	88.2
									17	90.1
									10	91.2
2	2								14	92.8
1	1	1							10	93.9
0	2			1					4	94.3
1	1	1							7	95.1
1	2	1	1						8	96.0
3		3	5	4	2				18	98.0
	3		2	1					6	98.7
				1	1				2	98.9
			1		1	0		2	4	99.3
			1		1		3		5	99.9
								1	1	100.0
8	11	6	10	7	5	0	3	3	898	
95.0	96.2	96.9	98.0	98.8	99.3	99.3	99.7	100.0		

dard deviation, which is also seen in Table 2. The values are therefore ln transformed. In Fig. 2A the differences of ln values are plotted against the laboratory method and in Fig. 2B against the average of the two measurements. In the ln-transformed figures, the 95% prediction interval and the between-method differences are nearly constant across the range of measurements. Of the data points, 8.2% (n = 74) are found outside the 95% prediction interval in the normal difference plot (Fig. 1B) and 8.8% (n = 79) in the ln-transformed plot. Both in the normal plot and in the ln-transformed plot the number of results outside the prediction interval and the number of results with a high deviation from zero are higher for the lower values of CRP.

The lines for the upper and lower limits of the 95% prediction interval are not perfectly linear (Figs. 1 and 2). This is because the accumulated CV for the two methods is not constant, but varies from 16% to 21% depending on the value of the CRP (see *Appendix* for the CV calculations). In Fig. 1A, all differences are positive when the laboratory CRP is 10 mg/L. This is because the lowest possible result for the near-patient test is <10 mg/L, which is defined as 9 mg/L in the calculations, and then a difference <-1 mg/L is not possible. In Fig. 1B this effect is balanced.

Discussion

The results obtained by the near-patient test are consistent with the results obtained at the laboratory. We found an

overall mean difference (95% CI) of -0.3 mg/L (-0.9 to 0.3 mg/L), indicating no major systematic bias between the two methods. The bias (mean of differences) demonstrated in all five intervals of CRP values (<10, 10-25, 26-50, 51-100, and >100 mg/L) was within ± 5.0 mg/L, indicating the near-patient test to be accurate for high as well as for low values of CRP. Regarding precision, ~8.5% of all data points (between-method differences) were found outside the 95% prediction interval (Figs. 1 and 2). We would expect ~5% outside this interval because of random variation only, and the surplus of ~3.5% of the data points must be explained by conditions not taken into account when the 95% prediction interval was calculated. Regarding values >10 mg/L by the laboratory method, the percentages of points outside the prediction interval is 20.4% (54 points of 264 possible), with the majority of points outside the limits between 10 and 50 mg/L. A reason for this could be the inevitable component of variation, comparing a whole-blood method with a serum method (i.e., because of lack of a correction for high or low hematocrit values), but such a component of variation would probably give an even variation independent of the concentration of the CRP. In our data a higher degree of imprecision is seen for the lowest values of CRP (Figs. 1 and 2), and a more probable explanation would be that it is more difficult to discriminate the lower values of CRP than the higher when using the near-patient test. This explanation seems reasonable, as the color intensity is low for lower values of CRP and thereby more difficult

Table 2. Mean difference between measurements of CRP when measured with a near-patient test in GPCs and measured with a turbidimetric method in the laboratory.

Overall results for all 13 GPCs					
GPC	No. of measurements	Mean of difference, ^a mg/L	95% CI of mean difference, ^b mg/L	SD of difference, ^c mg/L	Mean value laboratory, mg/L
1	18	0.5	-1.4 to 2.4	3.8	11.8
2	35	0.2	-1.6 to 2.0	5.2	13.4
3	23	-0.3	-1.4 to 0.8	2.3	14.4
4	75	-0.2	-1.9 to 1.4	7.1	17.6
5	47	-0.8	-3.4 to 1.9	8.9	17.6
6	241	0.7	0.4 to 1.8	8.5	19.0
7	24	-1.8	-5.3 to 1.6	8.1	20.5
8	25	0.6	-2.0 to 3.2	6.3	22.2
9	200	-2.1	-3.7 to -0.6	11.0	22.4
10	39	0.5	-2.5 to 3.4	9.1	23.1
11	86	0.3	-2.8 to 3.4	14.4	23.4
12	51	0.8	-1.2 to 2.8	7.0	25.5
13	34	-0.4	-2.0 to 1.3	4.6	25.6
1-13	898	-0.3	-0.9 to 0.3	9.6	20.3
Compared results of CRP measurements for five intervals of CRP values					
Laboratory values of CRP, mg/L					95% CI of SD ^d mg/L
<10	594	0.6	0.4 to 0.7	2.1	2.0 to 2.2
10-25	145	-0.7	-1.7 to 0.4	6.4	5.7 to 7.1
26-50	74	-4.5	-7.8 to -1.2	14.1	11.8 to 16.4
51-100	58	-1.5	-6.2 to 3.1	17.8	14.6 to 21.0
>100	27	-2.5	-17.4 to 12.4	37.7	27.7 to 47.7
Compared results of CRP measurements with GPCs dichotomized according to whether the near-patient test was operated by a technician or operated by unskilled laboratory personnel					
Personnel					
Technicians	398	0.5	-0.3 to 1.3	8.0	7.4 to 8.6
Nontechnicians ^e	500	-0.9	-1.8 to 0.0	10.6	9.7 to 11.3
Compared results of CRP measurements dichotomized in measurements obtained from all 13 GPCs within the first week or within the first or second week of the study period compared with measurements obtain in rest of period					
Study period					
First week	134	-1.5	-2.8 to -0.2	7.5	6.6 to 8.4
Rest	764	-0.1	-0.8 to 0.6	9.8	9.3 to 10.3
First and second week	301	-0.1	-1.2 to 1.0	10.0	9.2 to 10.8
Rest	597	-0.4	-1.1 to 0.4	9.3	8.8 to 10.1
Compared results of CRP measurements with GPCs dichotomized according to their number of performed tests in the study period					
No. of measurements per GPC					
<50	245	-0.2	-1.1 to 0.8	7.7	7.0 to 8.4
>50	653	-0.3	-1.1 to -0.5	10.2	9.6 to 10.8

^a Results given as mean difference between measurements of CRP with a near-patient test in GPCs and measurements of CRP with a turbidimetric method in laboratory (CRP measured with a near-patient test - CRP measured turbidimetrically).

^b 95% confidence interval for the mean difference.

^c Standard deviation for the differences between measurements of CRP in general practice and in laboratory.

^d 95% CI for the standard deviation for the differences.

^e Nurses, secretary staff, or the GPs themselves.

to evaluate precisely with visual evaluation. For diagnostic use, however, CRP values up to 20 mg/L are of uncertain diagnostic value, CRP values from 20-50 mg/L

are often associated with viral infections, and values >50 mg/L are usually due to bacterial infections [10, 11, 14]. So for the diagnostic purpose of distinguishing between

viral and bacterial infection, the larger imprecision for the lower values of CRP seems not as important as a good analytical precision for CRP values ~ 50 mg/L, which seems to be achieved by this near-patient test. It is possible that a better precision for the lower values could be achieved by using a color densitometer, but it is a question of whether it is of clinical relevance.

As in other studies from general practice [9, 14, 15], a majority of measured CRP values was found within the normal reference interval (i.e., <10 mg/L). Still, 20% ($n = 176$) of all test results were >20 mg/L, and 10% ($n = 92$) were high (≥ 50 mg/L). The large sample size makes evaluation of higher values of CRP possible. The percentage is in accordance with other studies in this field where 21% and 22% of all CRP measurements were >20 mg/L [14, 15], but the total number is higher because of the considerable sample size in the present investigation. The percentage of higher CRP values (≥ 50 mg/L) in our study population is nearly the same as in the other studies, indicating some homogeneity in the study populations and in the threshold to do the test among GPs [14, 15].

In a laboratory evaluation of the test comprising 230 paired measurements performed by experienced technicians and evaluated with a color densitometer, the results were grouped in the following intervals: 1–10, 11–25, 26–50, 51–100, 101–200, and >200 mg/L [13]. Of the results of both methods, 79.5% were found in the same groups and 99.6% were found within the same group or an adjacent group. If we make the same rough grouping from the data shown in Table 1, we find 92.6% within the same group and 99.6% within the same or an adjacent group, indicating that the test performance made in general practice is as good as obtained by experienced technicians performing the test in a laboratory.

Studies comparing two analytical methods (procedures) for measuring the same component often report results as either a correlation coefficient or as the number of "false-normal" and "false-elevated" values when compared with a reference method. These two test statistics are easy to use but not always informative. A correlation coefficient measures the relation between two variables (and not necessarily the agreement between the two variables) and is dependent on the range of the measured quantity [18]. If testing for significance is made, a high level of significance will nearly always be found, but it would otherwise be a surprise if two methods designed to measure the same quantity were not related [18].

NycoCard CRP is available for whole blood as well as for serum or plasma. In 1991 a Norwegian group tested the NycoCard CRP Serum/Plasma test kit in 194 samples and reported their results as $r = 0.85$ when the test was operated in 10 primary healthcare centers [14]. We found an overall r of 0.94. This difference may be due to fewer analytical steps in the whole-blood test than in the serum test, reflecting that few analytical steps improve the analytical quality. But the difference can also be explained by the fact that more data points for high values of CRP

also gives a higher correlation, as we found 27 CRP measurements >100 mg/L compared with 9–10 CRP measurements in the Norwegian study.

In a study from The Netherlands, NycoCard CRP Whole Blood was evaluated in general practice comparing 439 samples [15]. In 88% of all measurements, corresponding results were found between a laboratory method and the test, with CRP results at 25 mg/L as the cutoff value in general practice and CRP results at 20 mg/L as the cutoff value for the laboratory. Furthermore, the results were reported as the frequency of "false normal" and "false elevated," which were 8% and 28% respectively. The study group concluded: "The reliability of the NycoCard CRP measurement in whole blood is disappointing. In particular the 'false elevated' rate is unacceptably high . . ." In general, it is necessary to choose the same cutoff value for both methods to avoid misinterpretation of results when a cutoff value is used as evaluation strategy. Using data from Table 1 for reporting our results as the frequency of "false normal" and "false elevated," we found 2.9% and 6.6% respectively with a cutoff value of 25 mg/L and 1.2% and 4.6% respectively with a more clinically meaningful cutoff value of 50 mg/L.

When a test evaluation is made by comparing the test results with results from a laboratory, the evaluation is entirely dependent on the analytical quality of the laboratory. External quality assessments among clinical chemical laboratories have demonstrated major interlaboratory variation in CRP measurements. In a Belgian study the interlaboratory CV was between 19% and 23% when mailing two samples for CRP measurements (median values of 17.0 mg/L and 40.8 mg/L) to 345 laboratories [16]. So if a laboratory measures CRP too high or too low, it will have a major influence when using these results as "true values" in a test evaluation, especially if the evaluation is made only as a cutoff evaluation. The laboratory must be able to document its analytical results to be traceable to international reference preparations (i.e., BCR/CAP/IFCC CRM 470). In the two mentioned studies evaluating CRP measurement in general practice, no information is given ensuring the traceability of the laboratory results [14, 15].

A near-patient test in general practice will be operated by nurses, secretary staff, or by GPs themselves, and not so often by technicians. The robustness of the near-patient test is therefore an important factor to evaluate [3]. When comparing results from GPCs served by technicians with results from GPCs without technicians, we found no significant differences in the test quality regarding the bias (mean of difference) but a slight difference regarding the precision (SD of the differences) (Table 2). In the Norwegian study all serum samples were frozen and retested by one technician at the laboratory, and r then increased from 0.85 to 0.95 [14]. The Norwegian study group considered this effect to be due to an elimination of the interpersonal variation as well as a difference in the analytical experiences between the personnel in general

practice and the technician. Comparing these results with our results, it seems likely that the difference is mostly due to elimination of the interpersonal variation.

Another critical demand for a near-patient test is the time needed from the introduction of the test until the staff members are experienced as operators. In the study from The Netherlands the staff in general practice was given 2 weeks to practice CRP measurement before the evaluation was performed [15]. In our study the staff in general practice was given one introductory visit, but no time for practicing the test in advance of the study, as we wanted to evaluate the analytical quality during the learning period. We found no differences in the test results obtained within the first or second week compared with test results from the rest of the period, except for a slightly lower SD of difference in the first week of the study (Table 2). When comparing clinics with a high and clinics with a low request frequency (more or less than 50 tests in the 2-month study period), we found no differences regarding the mean of difference and only a small difference in the SD in favor of the clinics with a low request frequency (Table 2). These results indicate that the use of the test is not technically demanding and the test can be operated sufficiently after just one introductory visit.

APPLICATION OF DIFFERENCE PLOTS IN TEST EVALUATION

Bland and Altman have suggested using a difference plot when evaluating a new method with a standard method, and they recommend that the between-method difference be plotted against the average value of the two methods, thus avoiding a negative correlation between the differences and the standard method [18, 19]. In Fig. 1A we have plotted the between-method difference against the laboratory method. Calculating the expected negative correlation between the differences and the laboratory method by using the formula given by Bland and Altman [19], we find $r = -0.19$. The correlation between the actually measured differences and results of the laboratory method was $r = -0.16$. Making the same calculations between the differences and the average results of the two methods, we find an expected $r = 0.01$ and a measured $r = 0.01$. So, plotting the differences against the laboratory method gives a slightly negative correlation. In our test evaluation, we compare a bedside method with a known high variation ($CV_{\text{test}} = 15\text{--}20\%$) with a laboratory method with a known very low analytical imprecision ($CV_{\text{lab}} = 5\%$) (see Appendix for CV calculations). If the between-method differences are plotted with the laboratory method as the abscissa, the uncertainty because of known analytical variation for the abscissa will be low compared with a plot with the average of the two methods as the abscissa. This fact is illustrated in Fig. 3. In the Figure, boxes for the 95% analytical uncertainty of CRP values of 20, 50, 100, and 200 mg/L are shown. The height of the boxes illustrates the 95% analytical uncertainty for

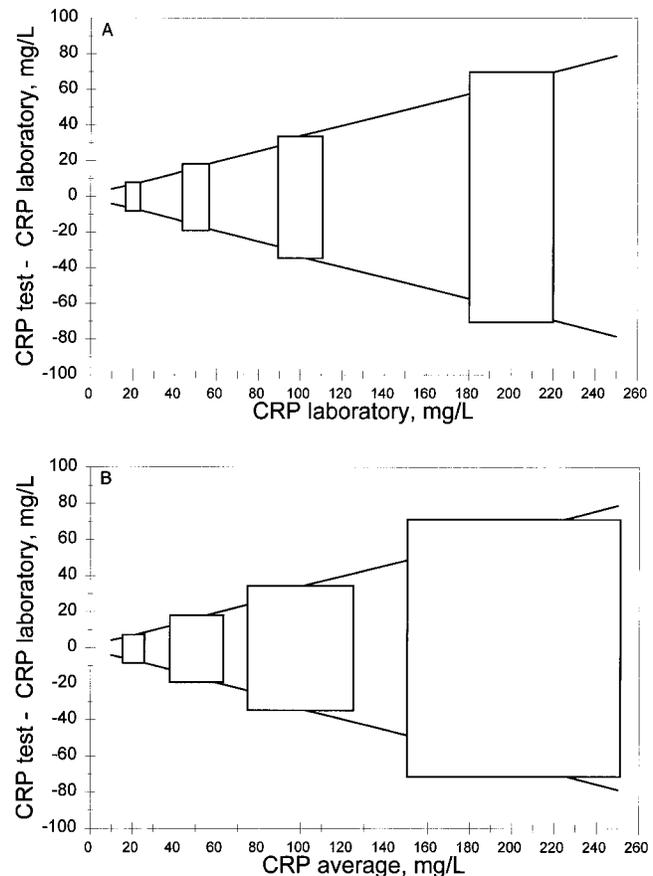


Fig. 3. The 95% analytical uncertainty calculated for CRP values of 20, 50, 100, and 200 mg/L.

The height of the boxes illustrates the 95% analytical uncertainty for the difference between the two methods, which is the same in both Figures. The width of the boxes illustrates the 95% analytical uncertainty of the method(s) plotted as the abscissa. (A) Uncertainty related to the abscissa when using the laboratory measurement is low; (B) uncertainty related to the abscissa when using average of the two measurements is high. The uncertainty on the abscissa is much larger when using the average of the two measurements, as the uncertainty is not only related to the low CV of the reference method, but to the average of the low CV of the reference method and the high CV of the near-patient test method (see text and Discussion for further details).

the difference of zero between the two methods. The width of the boxes illustrates the 95% analytical uncertainty of the abscissa. In Fig. 3A the laboratory method is plotted as the abscissa and in Fig. 3B the average of the two methods is plotted as the abscissa.

In Fig. 3A and B the 95% analytical uncertainty of the differences is calculated as:

$$0 \pm 1.96 \times \mu \times (CV_{\text{lab}}^2 + CV_{\text{test}}^2)^{1/2}$$

where μ is the value of CRP.

Example: With a true CRP value of 20 mg/L and a measured between-method difference of zero, the uncertainty of the difference is:

$$0 \pm 1.96 \times 20 \text{ mg/L} \times (0.05^2 + 0.17^2)^{1/2} \Rightarrow 0 \pm 6.9 \text{ mg/L}$$

In Fig. 3A, the 95% analytical uncertainty with the laboratory method plotted as the abscissa is calculated as:

$$\mu \pm 1.96 \times \mu \times CV_{\text{lab}}$$

Example: With a true CRP value of 20 mg/L, the uncertainty of the value is:

$$20 \text{ mg/L} \pm 1.96 \times 20 \text{ mg/L} \times 0.05$$

$$\Rightarrow 20 \text{ mg/L} \pm 1.96 \text{ mg/L}$$

In Fig. 3B, the 95% analytical uncertainty with the average measurements plotted as the abscissa is calculated as:

$$\mu \pm 1.96 \times \mu \times [(CV_{\text{lab}}^2 + CV_{\text{test}}^2)/2]^{1/2}$$

Example: With a true CRP value of 20 mg/L, the uncertainty of the value is:

$$20 \text{ mg/L} \pm 1.96 \times 20 \text{ mg/L} \times [(0.05^2 + 0.17^2)/2]^{1/2}$$

$$\Rightarrow 20 \text{ mg/L} \pm 4.91 \text{ mg/L}$$

In Fig. 3 and from the example with a CRP value of 20 mg/L it is clearly seen that the 95% uncertainty on the abscissa is much larger with the average measurements (Fig. 3B) than with the laboratory method (Fig. 3A). Thus, evaluating a method with a known high variation by comparing with a reference method with a known low variation and plotting the between-method difference against the reference method, measurements will give less uncertainty on the abscissa (but a slight negative correlation) than plotted against the average measurement. In conclusion, this means that both plotting against the reference method and the average of the two methods have some disadvantages.

Bland and Altman suggest making the 95% limits of agreement on the basis of plus or minus 2 SDs of the differences (or more precisely 1.96) with the limits based upon the results of the test evaluation [18, 19]. Thus, a 95% limit of agreements will show the interval where 95% of all data points always will be within when the distribution is gaussian. If the agreement between the two tests is high, the interval will be narrow; if the agreement is low, the interval will be wide.

Instead, we have tried to predict the interval within which we would expect 95% of the data points to be found, having knowledge about the variances (CV) of the two test methods. In other words: Before actually performing the practical part of the study, we are predicting a kind of acceptance limits for the test evaluation, according to the known impressions of the two methods obtained during professional testing in the laboratory. Thus, in this study the 95% prediction interval expresses the best obtainable results possible when the test is operated by staff in general practice not trained as analysts.

GENERAL APPLICATION OF EVALUATIONS OF NEAR-PATIENT TESTS

When evaluating a near-patient test it is important to ensure that the measurements are performed under circumstances as near daily routine as possible, and when

planning the study to ensure that the sample size is sufficient for an evaluation of all intervals of test results. Moreover, the analytical quality at the reference laboratory is of major importance to the result of the evaluation. Comparison of semiquantitative methods with quantitative methods gives statistical problems that demand some reservations, some of which are set out in this paper. We have tried to fulfill these major assumptions in this evaluation by:

- Making the evaluation in the field of different GPCs, some with technician staff and others with nurses, secretary staff, or GPs themselves operating the tests.
- Not introducing a training period, but as in "real life" start performing measurements just after the introductory visit from the manufacturer.
- Having a patient population sample size large enough to ensure validity for test results above the normal reference interval.
- Having a reference laboratory that has values traceable to an international reference preparation (i.e., BCR/CAP/IFCC CRM 470).
- Developing a statistical-graphical method, making semiquantitative methods comparable with quantitative methods.

NycoCard CRP for whole blood with visual evaluation of test results is as good when used in GPCs as could be expected from laboratory testing in terms of bias and precision as a semiquantitative test for near-patient testing. The technical performance is nearly equal for technician and nontechnician staff, and for clinics with frequent and nonfrequent use of the test kit, indicating that use of the kit is not technically demanding. Before introducing the test in daily routine in general practice, we recommend an evaluation of the clinical effectiveness of introducing the test, as introduction of a new test should be determined by medical needs rather than by availability [21].

Appendix

CALCULATION OF THE 95% PREDICTION INTERVAL

When comparing two analytical methods (procedures) for measuring one and the same component by use of patient samples, both measurements are exposed to some variation, and the differences should be distributed around zero randomly with a predictable uncertainty. To predict the uncertainty, the split-sample technique makes it possible to avoid biological variation and sampling variation. The expected variance of differences is reduced to the sum of variances according to analytical imprecision of the two methods plus a component from preanalytical handling of the samples. As CRP is stable for several days, and the effect of the mailing procedure is considered negligible, then the expected uncertainty is related to the imprecision of the two methods only. The expected total CV (CV_{total}) can be estimated as:

$$CV_{\text{total}} = (CV_{\text{lab}}^2 + CV_{\text{test}}^2)^{1/2} = (CV_{\text{lab}}^2 + CV_{\text{test-analytical}}^2 + CV_{\text{test-read}}^2)^{1/2}$$

CV_{lab} , the coefficient of total analytical variation for the laboratory method, is 5% (see *Materials and Methods*).

$CV_{\text{test-analytical}}$ is the CV for the near patient test, $CV_{\text{test}} = 15\%$, approximated from the laboratory study [13] where results were performed by 11 experienced technicians and the reading of the color response was measured with a color densitometer (see *Materials and Methods*).

$CV_{\text{test-read}}$ is the CV for the readings. Using semiquantitative tests, reporting the test results in intervals, we normally have no specific knowledge about a possible value within an interval, and we can assume the value to lie anywhere within the interval, which gives us a rectangular distribution. Results from general practice were reported as "best guess" (within 5 mg/L from 10–50 mg/L, within 10 mg/L from 50–100 mg/L, and within 25 mg/L for values >100 mg/L). Thus, if a "true value" of CRP is within the interval of 17.5–22.5 mg/L, the midpoint answer 20 mg/L is considered correct for all CRP values within this interval, introducing a reading variation. This reading variation can be estimated by using the formula describing the distribution of uncertain values within a rectangle [20]:

$$CV_{\text{test-read}} = a / (2 \times 3^{1/2} \times \mu)$$

where a is the range between the two outer limits for the "true value" interval ($a = 5$ mg/L for CRP value from 10 to 50 mg/L, 10 mg/L for values from 50 to 100 mg/L, and 25 mg/L for values >100 mg/L) and μ is the midpoint answer for the interval. This means that the $CV_{\text{test-read}}$ is not constant, but varies according to the quantities of the measurements (μ):

Example 1A. At a CRP value of 20 mg/L, $\mu = 20$ mg/L and $a = 5$ mg/L:

$$CV_{\text{test-read}} = [5 / (2 \times 3^{1/2} \times 20)] \times 100 = 7\%$$

Example 2A. At a CRP value of 175 mg/L, $\mu = 175$ mg/L and $a = 25$ mg/L:

$$CV_{\text{test-read}} = [25 / (2 \times 3^{1/2} \times 175)] \times 100 = 4\%$$

Knowing the $CV_{\text{test-read}}$ it is possible to calculate the CV_{total} for different quantities of CRP:

Example 1B. At a CRP value of 20 mg/L:

$$CV_{\text{total}} = (CV_{\text{lab}}^2 + CV_{\text{test-analytical}}^2 + CV_{\text{test-read}}^2)^{1/2} = (0.05^2 + 0.15^2 + 0.07^2)^{1/2} \times 100 = 17\%$$

Example 2B. At a CRP value of 175 mg/L:

$$CV_{\text{total}} = (0.05^2 + 0.15^2 + 0.04^2)^{1/2} \times 100 = 16\%$$

The upper and lower limit for the 95% prediction interval can be calculated as:

$$0 \pm 1.96 \times CV_{\text{total}} \times \mu$$

As it appears from calculations above, the CV_{total} differs for different concentrations of CRP. In the CRP interval from 10 to 250 mg/L, CV_{total} varies from 16% to 21%, which explains why the lines for 95% prediction intervals are not perfectly linear.

CALCULATION OF THE 95% PREDICTION INTERVAL WHEN THE DIFFERENCES ARE LN TRANSFORMED
A ln-transformed CV is calculated as [22]:

$$CV = \{\exp[s^2(\ln)] - 1\}^{1/2}$$

where s is standard deviation of the logarithmic values. Solving the formula gives:

$$s(\ln) = [\ln(CV^2 + 1)]^{1/2}$$

and the upper and lower limit for the 95% prediction interval can be calculated as:

$$\pm 1.96s(\ln)$$

Example 1c. At a CRP value of 20 mg/L, $CV_{\text{total}} = 17\%$:

$$s(\ln) = [\ln(0.17^2 + 1)]^{1/2} = 0.17$$

and limits for the 95% prediction interval are $\pm 1.96 \times 0.17 = \pm 0.33$.

Example 2c. At a CRP value of 175 mg/L, $CV_{\text{total}} = 16\%$:

$$s(\ln) = [\ln(0.16^2 + 1)]^{1/2} = 0.16$$

and limits for the 95% prediction interval are $\pm 1.96 \times 0.16 = \pm 0.31$.

We thank all participating general practitioners and their assistants, as well as the laboratory staff of Vejle County Central Hospital. This work is supported by The Fund for Medical Research in Vejle County (j.nr:13/1995), The Danish Health Insurance Fund (j.nr:11/024–96), The Danish Research Foundation for General Practice (j.nr:FF-2-02-16), and The Memorial Foundation for Johs. M. Klein and wife (J.nr: J 664.14). Nycomed Pharma AS has sponsored all near-patient tests and the Danish division has introduced the GPs and their assistants to the near-patient test.

References

1. Hobbs R. Near patient testing in primary care [Editorial]. *BMJ* 1996;312:263–4.
2. Woolf S, Kamerowdb H. Testing for uncommon conditions. The heroic search for positive test results. *Arch Intern Med* 1990;150: 2451–8.
3. Association of Clinical Biochemists. Guidelines for implementa-

- tion of near-patient testing. London: Royal Society of Chemistry, 1993:8pp.
4. Dybkær R. Vocabulary for describing the metrological quality of a measurement procedure. *Ups J Med Sci* 1993;98:445–86.
 5. Soderstrom M, Blomberg J, Christensen P, Hovelius B. Erythromycin and phenoxymethylpenicillin (penicillin V) in the treatment of respiratory tract infections as related to microbiological findings and serum C-reactive protein. *Scand J Infect Dis* 1991;23:347–54.
 6. Stuart J, Lewis SM. Monitoring the acute phase response [Editorial]. *BMJ* 1988;297:1143–24.
 7. Young B, Gleeson M, Cripps AW. C-reactive protein: a critical review. *Pathology* 1991;23:118–24.
 8. Thompson D, Milford Ward A, Whicher JT. The value of acute phase protein measurements in clinical practice. *Ann Clin Biochem* 1992;29:123–31.
 9. Dahler-Eriksen BS, Lassen JF, Lund ED, Lauritzen T, Brandslund I. C-reactive protein in general practice—how commonly is it used and why. *Scand J Prim Health Care* 1997;15:35–8.
 10. Korppi M, Kröger L. C-reactive protein in viral and bacterial respiratory infection in children. *Scand J Infect Dis* 1992;25:207–13.
 11. Lindbäck S, Hellgren U, Julander I, Hansson L-O. The value of C-reactive protein as a marker of bacterial infection in patients with septicaemia/endocarditis and influenza. *Scand J Infect Dis* 1989;21:543–9.
 12. Philip AG, Andrews PA. Rapid determination of C-reactive protein levels: semiquantitative versus quantitative. *J Pediatr* 1987;110:263–5.
 13. Urdal P, Borch SM, Landaas S, Krutnes MB, Gogstad GO, Hjortdahl P. Rapid immunometric measurement of C-reactive protein in whole blood. *Clin Chem* 1992;38:580–4.
 14. Hjortdahl P, Landaas S, Urdal P, Steinbakk M, Fuglerud P, Nygaard B. C-reactive protein: a new rapid assay for managing infectious disease in primary health care. *Scand J Prim Health Care* 1991;9:3–10.
 15. Dinant GJ, Costongs R, Leclercq RM, Van Wersch JW. Reliability of C-reactive protein measurement in general practice in The Netherlands. *Scand J Clin Lab Invest* 1994;54:113–7.
 16. Devleeschouwer N, Libeer JC, Chapelle JP, et al. Factors influencing between-laboratory variability of C-reactive protein results as evidenced by the Belgian External Quality Assessment (EQA) Scheme. *Scand J Clin Lab Invest* 1994;54:435–40.
 17. Baudner S, Bienvenu J, Blairup-Jensen S, et al. The certification of a matrix reference material for immunochemical measurement of 14 human serum proteins. Report EUR 15243 EN. Commission of the European Communities, 1993.
 18. Bland JM, Altman DG. Statistical methods for assessing agreement between two methods of clinical measurement. *Lancet* 1986;i:307–10.
 19. Bland JM, Altman DG. Comparing methods of measurements: why plotting difference against standard method is misleading. *Lancet* 1995;346:1085–7.
 20. International Organization for Standardization. Guide to the expression of uncertainty in measurement, 1st ed. Geneva, 1994: 101pp.
 21. Sandberg S. Near patient testing must improve patient care [Editorial]. *Scand J Prim Health Care* 1994;12:65–7.
 22. Shimizu K, Crow EL. History, genesis, and properties. In: Crow EL, Shimizu K, eds. Lognormal distributions, theory and applications. New York: Marcel Dekker, 1988:1–25.