J. chron. Dis. 1969, Vol. 22, pp. 203–215. Pergamon Press. Printed in Great Britain

THE INTRAFAMILIAL TRANSMISSION OF RHEUMATOID ARTHRITIS—II

AN INTERVIEW MEASURE OF RHEUMATOID ARTHRITIS

SIDNEY COBB, M.D., M.P.H., PATRICIA HUNT, B.A.
and ERNEST HARBURG, PH.D.

Survey Research Center, The University of Michigan, Ann Arbor, Michigan

(*Received* 21 *December* 1967; *in revised form* 26 *August* 1968; *further revised* 6 *January* 1969)

RHEUMATOID arthritis is a remittent disease, and in its milder forms the people involved spend more time free of the disease than they do with the disease [1, 2]. This means that at a single clinical examination those with mild disease are unlikely to be identified because there is a high probability that they will be free of disease at any given point in time. More specifically, those who spend on the average only 10 weeks per year with joint swelling will have only 1 chance in 5 of being classified as having rheumatoid arthritis at a single examination because the diagnostic criteria of the American Rheumatism Association [3] require that the swelling be observed by the physician. Thus, population studies using a single clinical examination identify a group of rheumatoid arthritics that is overloaded with persons who spend a large proportion of their time in an episode of active disease and that is underloaded with persons who spend only a small proportion of their time in episode. When one is trying to characterize persons with the disease in regard to some set of variables, it is important that one include the full range of severity and not confine one's study simply to severe and continuously affected cases.

Since repeated examination of a population sample, especially of a national sample, would be far too costly, and since it seemed highly unlikely that even if the money were available, one could hire a sufficient number of physicians to make the necessary examinations, a carefully programmed interview for repeated use has been developed and partially validated. This interview can be administered by any reasonably intelligent high school graduate who has had appropriate training in survey methods [4]. It turns out to be a useful tool for studying the association between rheumatoid arthritis and a variety of variables that can similarly be estimated by interview.

The following principles guided the development of this interview form:

(1) The interview must be fully specified leaving no room for choice or judgment on the part of the interviewer.

(2) The principal focus of the interview must be on the diagnostic criteria for rheumatoid arthritis established by the American Rheumatism Association.
(3) It is desirable to have some estimates of activity and disability in order to judge severity of the disease, because the assurance with which one can diagnose this disease is highly correlated with the severity of the disease.
(4) The interview must seem brief and simple to those free of rheumatoid arthritis, but it is assumed that those who have arthritis are quite willing to talk about their diseases at some length so a greater number of questions directed at them is no problem.
(5) As indicated by RUBIN *et al.* [5], high specificity is more important than high sensitivity.

THE MEASURE

The form* is 7 pages in length and contains approximately 40 separate items, all of which are to be asked exactly as written. Of these 40 items only 12 are answered by those who are free of any arthritis. This means that a negative interview can be completed in 3–5 min., while an interview with a person having severe long-standing rheumatoid arthritis may run as long as 20 min. There were only 2 areas in which there was real difficulty in developing this form, and both were related to problems of exclusion. The first problem was ankle edema, and it took considerable experimentation to develop an appropriate set of questions which would exclude swollen ankles due to edema and yet include ankle swelling if it was surely arthritic. It should be recognized that it is sometimes quite difficult for a physician to recognize joint inflammation in the ankle when edema is also present and, in fact, it commonly takes more than 1 examination to be sure what the situation is. The other problem that proved somewhat troublesome was the development of an adequate technique for exclusion of persons with certain other diagnoses. On the one hand it was found that most people who had had a diagnosis of another disease that resembled rheumatoid arthritis knew this and were quite able to respond to the name of the disease. On the other hand, it is quite hopeless to try to exclude people with unrecognized collagen diseases, if they had never been diagnosed. However, because of the rarity of these diseases, very little bias is introduced by the failure to exclude them. There was some difficulty with people being excluded because several diagnoses had been considered by their physicians. We felt this was not a matter for serious concern because this group included those cases in which the diagnosis remained in doubt, and we were glad to have the doubtful cases excluded. Probably the greatest difficulty came in trying to exclude ankylosing spondylitis with peripheral involvement. Due to our exclusion of most cases for which pain was worst in the back, we have surely excluded some male rheumatoids with back pain.

From the questions asked, 3 indices and a basis for exclusion were constructed. The first, the Security of Diagnosis, was constructed by counting the number of the relevant manifestations that were reported as present at some point in any one of

*Copies of this form may be obtained by ordering NAPS Document 00294 from ASIS National Auxiliary Publications Service, c/o CCM Information Sciences, Inc., 22 West 34th Street, New York, New York 10001; remitting $1.00 for microfiche or $3.00 for photocopies.

the three interviews. The list adheres as far as is possible to the diagnostic criteria of the American Rheumatism Association. Obviously, the laboratory aspects of the criteria are irrevelant and some adjustments had to be made for the way people answer the specified questions. For example, a history of joint pain turned out to be too common and non-specific to be useful in discriminating rheumatoids from others. The seven manifestations of this index are as follows:

(1) Morning stiffness of 10 or more minutes duration. This item is more rigid than the A.R.A. criterion in accordance with the fifth principle above, i.e. specificity is more important than sensitivity.

(2) Joint swelling after correction for edema. The correction of ankle and foot swelling for edema was done by excluding those persons who met the following criteria:
   (*a*) Bi-lateral simultaneous report of swelling in both feet and both ankles that is worse in the afternoon or evening.
   (*b*) Essentially painless swelling that goes up the leg above the ankle.
   (*c*) Any ankle swelling thought by the respondent not to be due to arthritis.
   (*d*) Any mention of varicose veins or edema in marginal comments on the interview. It now seems likely that this last criterion could be improved by including a couple of appropriate questions on the subject of varicose veins and peripheral edema.

   Swelling in distal interphalangeal joints was not included. There is surely both under- and over-reporting of joint swelling; but there seems to be a slight tendency for the reporting to exceed that observed [6].

(3) Two or more joints swollen after correction for edema. Again, the terminal interphalangeal joints are excluded but all other peripheral joints are counted.

(4) Symmetrical joint swelling, which means simultaneous involvement of symmetrical joints, except that bilateral involvement of midphalangeal, metacarpophalangeal or metatarsophalangeal joints is accepted without absolute symmetry, as in the A.R.A. criteria.

(5) Duration of swelling is 6 weeks or more. This item is inserted as one of the manifestations. This is different from the A.R.A. criteria which say that definite rheumatoid arthritis is characterized by at least 5 of the criteria plus a duration of symptoms of at least 6 weeks. We had intended to adhere to the thinking of the A.R.A. criteria, but the duration of symptoms question that was used appeared to by rather non-specific and to produce answers that were inconsistent with other data in the interview. For this reason we turned to the question on duration of swelling. It was obvious that it would be too exclusive to require a report of at least 6 weeks of continuous joint swelling when the A.R.A. required only 6 weeks of continuous symptoms. Therefore, the decision was made to use the duration of swelling as one of the criteria, rather than as an exclusion.

(6) Report of a physician diagnosis of rheumatoid arthritis. Since we were using only historical data, it seemed appropriate to use the available information about physician diagnosis. It turns out to relate reasonably well to the other criteria and to discriminate better than the history of elbow nodules.

(7) History of elbow noludes. These are highly specific for the true diagnosis, but are sufficiently rare so that they contribute nothing to the sensitivity of the measure.

The Activity Index was constructed by taking the first 4 of the following variables and bracketing them into 6 categories, i.e. zero plus 5 equal intervals. The fifth one was scored zero for no and 3 for yes. The mean was then taken and the results were re-bracketed into a similar scale from zero to 5:

(1) The sum of the number of joints swollen in all 3 interviews;

(2) The mean number of days with swelling for the 3 periods of 4 weeks about which inquiry was made;

(3) The mean duration of morning stiffness;

(4) The mean number of aspirin tablets taken per day for arthritis;

(5) A report of steroid and/or gold therapy.

The Disability Index was constructed by taking the mean of a disability score and the days unable to perform usual duties and bracketing to give a range from zero to three. The disability score was arrived at by taking the mean degree of severity for each of 10 disability items, such as going up or down stairs, stooping or bending and picking up or using small things, after correction for comments indicating that a particular disability was not really due to arthritis. The number

TABLE 1. THE RELATIONSHIP BETWEEN THE COMPONENTS OF THE RA MEASURE WITHOUT REGARD TO EXCLUSION

	Activity						Disability				Total
	0	1	2	3	4	5	0	1	2	3	
Security of *Dx*											
0	178	6					158	26	1		184
1	6	30					15	20	1		36
2	1	11					2	11	1		15
3		12	6	1			4	13	2		19
4		6	9	4			2	13	4		18
5		2	12	10	2		1	21	3	1	26
6				4	8	1	5	5	3	5	13
7			2	1	7	3		4	5	4	13
Total							181	113	20	10	324

							Total
Disability							
0	159	21	1				181
1	25	45	23	13	7		113
2	1	1	7	4	4	3	20
3				3	6	1	10
Total	185	67	31	20	17	4	324

of days unable to perform usual duties was simply the mean over the several interviews of the number of days out of the preceding four weeks for which the respondent reported that he was unable, because of arthritis or rheumatism to carry out all his/her work on the job or around the house.

The interrelationships among these three variables are shown in Table 1. These relationships do not consider the exclusions. The relationships are strong enough to suggest that it is appropriate to combine the indices.

The diagnosis of rheumatoid arthritis was excluded if the coders found that the respondent admitted to any of the following diagnoses on direct question or, rarely, mentioned them in marginal comments: rheumatic fever, rheumatic heart disease, psoriasis, scleroderma, dermatomyositis, systemic lupus erythematosus, gout, arthritis due to infection, Reiter's disease, spondylitis. Furthermore, those few cases reporting arthritis worst in the spine or with onset before age 13 were reviewed by the senior author and a decision made about exclusion. This means that appreciable numbers of true rheumatoids were excluded simply because one of the above diagnoses had been entertained. We feel confident that most of the cases in which the diagnosis was in doubt were identified by these questions and that though the exclusion was a little overdone, this was appropriate in line with principle number five, above, that specificity is more important than sensitivity.

At this point we have three separate variables to be combined into a single score for the RA measure:

1. Security of diagnosis ranging from 0 to 7.
2. Activity measure ranging from 0 to 5.
3. Disability measure ranging from 0 to 3.

If an exclusion is present, the score is automatically zero. Furthermore, in order to prevent an accumulation of points from activity and disability (i.e. complaints), alone leading to a positive RA measure, it was further required that unless 3 points were derived from the security of diagnosis, the score would automatically be zero. Hence all cases for which the security of diagnosis was 0–2 were coded zero on the RA measure. The sum of the three indices (7+5+3=15) provides a scale reaching 15. The final RA measure was bracketed as follows:

Sum of indices	RA measure
1–3 or exclusion	0
4	1
5	2
6	3
7	4
8	5
9–10	6
11–13	7
14–15	8

Scores of one or two were considered doubtful and usually bracketed with the zeros as not rheumatoid arthritis. Stated another way, in order to be positive on the RA measure a respondent had to report no exclusions, get at least 3 points from the security of diagnosis scale and 3 more points from any of the 3 scales.

It is important to note that the security of diagnosis was given the largest weight and the complaint related disability measure the smallest weight.

VALIDATION

The validation of this RA measure involved multiple interviews with 75 patients from the arthritis clinic of the University of Michigan Hospital, which is under the direction of Dr. Ivan Duff. Of these patients, 23 were derived from the study described in the first paper of this series, though only 21 of them were actually used in that study; and 52 were additional patients attending the clinic. The only apparent bias in the selection, other than those incident to clinic attendance, is that patients coming more frequently had a greater chance of being included in the sample because they were more readily available. The procedures for these 2 groups were not quite identical in that the first group of 23 were interviewed on 3 occasions 4 months apart, as were all the other persons in the main study. By the time the validation study got under way we had come to realize that a set of 3 reports about arthritis which might be even more useful could be obtained with only 2 interviews by inquiring on the first occasion with regard to the present and with regard to the worst attack and at the second interview just with regard to the present. As a result, the 52 additional patients were interviewed using this more efficient method which we would recommend to others who might wish to use this technique.

Each clinic record was read by the senior author and graded for diagnosis of rheumatoid arthritis by the A.R.A. criteria [3] and for severity by estimating activity, proportion of time in episode, and disability. These were summarized as follows:

0=no rheumatoid arthritis
1=possible or probable disease with minimal involvement
2=probable or definite disease with moderate involvement
3=definite disease with severe involvement
4=classical disease that is active and chronically incapacitating.

The comparison of this classification with the RA measure is shown in Table 2.

TABLE 2. THE RELATIONSHIP OF THE CLINIC FINDINGS TO THE RA MEASURE WITH AN ESTIMATE OF THE SENSITIVITY OF THE RA MEASURE

RA measure severity	Clinic chart grade				
	0–1	2	3	4	Total
0–2	14	1	2	5	22
3–6		6	5		11
7	4	3	13	4	24
8		1	7	10	18
Total	18	11	27	19	75
0–2			8		
3–8			49		
			57		

$$\text{Sensitivity} = \frac{49}{57} \times 100 = 86 \text{ per cent}$$

The reader will recall that the RA measure was established by coders following a specified set of instructions; therefore, the 2 classifications are quite independent. (In this sample there were no cases of pain worst in back or onset before age 13 that required adjudication.) The relationship is generally strong and the biggest errors of misclassification are exclusions by the interview of cases for which multiple diagnoses were considered in the course of the workup. All 8 of the false negatives are of this nature. When the categories are collapsed in the lower part of the table, it is possible to calculate the sensitivity of this measure, which is the percentage of those truly diseased who are positive on the measure. The obtained value of 86 per cent is quite satisfactory.

The differences between the first and second samples were trivial, supporting the notion that 2 interviews covering 3 time periods are quite as satisfactory as 3 interviews. On dividing the total sample by sex it was noted that the sensitivity estimates were not significantly different (men=77 per cent and women=90 per cent). The slight reduction in sensitivity in men is due to excess exclusions for questionable gout and spondylitis.

The estimation of specificity from clinic data is unsatisfactory because there are too few true negatives and the values always come out too low. We have, therefore, elected to estimate the specificity by what might be called the method of completing the fourfold table. This is analogous to the procedures suggested by CORNFIELD [6] and DORN [7]. Table 3 presents an artificial population of 1000 persons equally distributed between the sexes. First, the National Health Survey Data [8] were used

TABLE 3. ESTIMATION OF THE SPECIFICITY OF THE RA MEASURE USING AN ARTIFICIAL POPULATION

	"True" diagnosis		
RA measure	Not RA	RA	Total
0–2	898	12	910
3–8	14	76*	90†
Total	912	88‡	1000

$$\text{Specificity} = \frac{898}{912} \times 100 = 98 \text{ per cent.}$$

*Sensitivity estimated at 86 per cent, $88 \times 0.86 = 76$.
†Nine per cent of the cousins and unrelated persons in this study are positive on the RA measure.
‡From the Oak Ridge Studies [6] and the National Health Survey it is estimated that 8.8 per cent of persons 35 yr of age and older would be found to have rheumatoid arthritis after 3 screenings at least 3 months apart.

to get a national estimate of the point prevalance of RA in persons 35 yr of age and older. This is 4.7 per cent. Then looking at Table VIII of the BEAL and COBB report [9], one sees that the average increase in number of persons diagnosed rheumatoid arthritis after two additional examinations 3 and 6 months after the first is from 11.5 to 21.6. Applying this rate of increase to 4.7 per cent, we get 8.8 per cent as the estimated true prevelance of rheumatoid arthritis in the population 35 yr of age and older when 3 examinations are made. Next, we get the other marginal total from the current study by taking the cousins and the unrelated persons as a source for an estimate of the population frequency of a positive RA measure. This is 9

per cent or 90 out of the thousand, which is remarkably close to the true population estimate. With the marginal totals all we need is one cell filled to fix the numbers in the entire table. Taking the estimate of the sensitivity of the measure, 86 per cent from Table 2, and applying it to the 88 persons with 'true' rheumatoid arthritis, we get 76 persons correctly classified by the measure and 12 false negatives. After completing the table, we can calculate the specificity

$$\frac{898}{912}\times 100=98 \text{ per cent.}$$

The sensitivity of 86 per cent and specificity of 98 per cent can be combined into an index for evaluating tests that was devised by YOUDEN [10] (J=sensitivity+specificity -1). In this case $J=0.86+0.98-1=0.84$. This is substantially better than the RA index of RUBIN *et al.* [5] for which the sensitivity=66 per cent, the specificity =95 per cent and $J=0.61$.

If one were to look at Table 3 as an assessment of agreement between two methods, it would be more appropriate to use the symmetrical form of Youden's measure which works out to be

$$J_s=\frac{1}{2}\left(\frac{a}{a+b}+\frac{c}{a+c}+\frac{d}{b+d}+\frac{d}{c+d}-2\right)$$

when the fourfold table is made up

a	b
c	d

This is related to the A_1 of ROGOT and GOLDBERG [11]

$$J_s=\frac{4A_1-2}{2}$$

and has the desirable property of ranging from -1 to $+1$ like a correlation coefficient instead of from 0 to 1 as the A_1 of Rogot and Goldberg. Since Table 3 is very nearly symmetrical, J_s is also 0.84. This compares reasonably with that of 0.86 for the agreement of 2 observers assessing byssinosis in the same population [12] and compares very favorably with two observers assessing coronary heart disease 0.50 [13] and hypertensive heart disease 0.61 [14] on the same individuals. It is also better than the agreement between radiologists with respect to the diagnosis of tuberculosis 0.66 [15], osteoarthritis of distal interphalangeal joints 0.56 [16] and peptic ulcer 0.48 [17].

This rather satisfactory result should not be looked on as the last word in validation. For example, there are other methods of doing the computations such as the procedure of mixed group validation suggested by DAWES and MEEHL [18]. Using this procedure, with the validation group and the national data as the two groups, one gets estimates of sensitivity of 93 per cent and specificity of 99 per cent. However, we are inclined to think that their assumptions are not really fulfilled in that the validation group is not a proper population and as such gives a poor specificity estimate.

Subsequent users of the RA measure should take advantage of minor improvements and simplifications that we can suggest and should undertake further validation studies. It is important to remember that a single clinical examination is not a

sufficient basis for validation because the RA measure picks up additional cases at each of the interviews. In fact, it is not unlikely that the RA measure comes closer to 'God's Opinion' with regard to who has rheumatoid arthritis and who does not than does a *single* clinical examination. For example, as noted above, a single examination has a sensitivity of only

$$\frac{4.7}{8.8} \times 100 = 54 \text{ per cent (page 11)}$$

for the cases that would be identified on a series of three examinations.

ILLUSTRATIVE CASES

The following cases have been selected in order to indicate the nature and extent of the information available from the interviews and to illustrate the types of cases that fit into the several severity categories.

(1) This is a 55-yr old man who has had rheumatoid arthritis since the age of 38. He is a former company executive, then owner of a dry cleaning business who is now unable to work. His second wife supports the two of them with a job as head cashier in a department store at which she makes about $4000 per year.

At the age of 38 he had an extended period of joint swelling involving wrists, knees and ankles that forced his retirement from his executive position. This attack subsided and he was relatively free from his disease until the age of 53 when he developed what was called 'walking pneumonia'. Following this he had a reactivation of his arthritis which has not let up since.

In 1962 he was hospitalized for 19 days and in February of 1964 he was again hospitalized for about 3 months.

At each of the 3 interviews he reported swelling in nearly every joint in his body along with morning stiffness of at least one and one-half hours duration. He reported elbow nodules. At one point his knees required repeated taps. Most of the joints in his wrists, ankles and feet are said to be ankylosed. He reports that he regularly takes 7 or more grams of aspirin per day. He has extensive disabilities, including the characteristic difficulty of dropping things which implies a weak grip.

He does not report any diagnoses suggesting an exclusion and he is quite clear that his diagnosis is rheumatoid arthritis. In addition he reports a peptic ulcer with onset and radiologic diagnosis at about age 30.

His rating is 8. Since he is a key person in the clinic sample, we also have a clinic reading on him which was 4. This means that he had the highest possible rating in each of the evaluations.

(2) This is a 54 yr old married woman who has 2 adopted children. She was married at the age of 30 to a man 4 yr younger than herself who is now self-employed as a beer and beverage distributor.

She reports that she first had arthritis at about age 50 and that for the last 2 yr she has had continual joint pain and substantial episodes of joint swelling. At the time of the first interview she had morning stiffness lasting 4 hr with swelling in both knees and the left ankle. At the time of the second interview she had morning stiffness lasting 2 hr and reported swelling in the PIP 3 and MCP 3 joints on the right, in both knees and in both ankles; but she reported that she had only had 7 days of swollen joints during the preceding month. At the time of the last interview she reported morning stiffness lasting one and one-half hours, swelling in the right

MCP 2, both knees and the right ankle. At this point she reported that her joints had been continuoulsy swollen for the preceding 28 days and said that the longest period of swelling that she had had was 5–6 months. She also reported that though she did not go to a doctor much, a doctor had told her that he believed her disease was rheumatoid.

She is quite obese, weighing 200 lb with a height of 5ft 3in. She is also reported to have some hypertension for which she is taking hydrodiuril. She does not respond positively to any of the questions which might lead to an exclusion of the diagnosis.

Her rating is 6, indicating a well-defined case of rheumatoid arthritis.

(3) This 26-yr old married housewife is an identical twin, and since she was a few minutes older than her twin sister, she was born eighth in a family of nine. She attended high school and has had some business college training. At the age of 18 she married a structural iron worker who is now employed at a job at which he earns about $6000 a year. She and her husband are both members of the Mormon Church.

At the age of 17 she had her first episode of arthritis; this involved her fingers and interfered with her typing. It lasted about a month. At the first interview she complained of pain in her knees and swelling in her ankles. The ankle swelling probably was due to the fact that she was then pregnant. She volunteered the information that her arthritis is apt to be related to 'nervous crises'. She did not have any joint swelling reported at the second interview and she did not mention significant morning stiffness on either of these occasions.

At the time of the third interview she had severe morning stiffness and had extensive symmetrical involvement of PIP's and MCP's, wrists and knees. She reports that though the swelling had been present at the time for only 10 days, she had been suffering with joint pain continuously for the past 2 yr. Even though she describes substantial joint swelling, she does not describe herself as very much disabled. She says that her doctor has told her that she might have rheumatoid arthritis. She was not taking any medicines for her arthritis, though she said that she had obtained considerable relief from a prescription that her doctor gave her on a previous occasion.

In addition to the rheumatoid arthritis, she reports symptoms of a peptic ulcer and says that she has had some kind of thyroid trouble. The pregnancy which contributed to her joint swelling at the time of interview B terminated with a miscarriage and was followed by what she believes to be an 'infection in her tubes', which 4 months later was still not completely cleared up.

Her rating is 5. This means that her disease is rated as of only moderate degree. She is of particular interest for she would not have been detected as having rheumatoid arthritis had we conducted only 2 interviews.

(4) This 32-yr old mother of nine children left the tenth grade of school to get married at the age of 16. Her husband is now manager of an Iowa hog farm. They report a cash income of about $3000.

At the first 2 interviews she did not report arthritis, swollen joints or morning stiffness. At the third interview she admitted to intermittent morning stiffness and reported that both knees had been swollen for about a week. She did not report herself as appreciably disabled by this, nor was she taking any medication. Since she had not sought medical advice, she did not know the nature of her arthritis

and answered negatively with regard to the various questions about other diseases which might exclude the diagnosis of arthritis.

Her rating is 2. This kind of a case gets a minimal score on the arthritis scale and certainly is one about which one would have doubts from a clinical standpoint until it had been watched for a longer period of time. Such a case is considered doubtful and included with the zeros when separating those with a positive RA measure from those without.

Up to this point, we have dealt with a 9-point scale. This degree of detail has been useful in the validation but the requirements of the analyses to follow are not so stringent, so a four point scale will be used. This relates to the nine point scale as follows:

Four-Point Scale	Nine-Point Scale
0	0
1	1–2
2	3–5
3	6–8

Categories 2 and 3 on the 4-point scale are referred to as RA. The distribution of RA in the sample is laid out in Table 4.

TABLE 4. THE DISTRIBUTION OF PERSONS WITH RA IN THE SAMPLE

Key person has RA	Males	Females	Total
Key person	12	31	43
Key person's spouse	2	1	3
Sib	1	2	3
Sib's spouse	0	5	5
Paternal cousin	0	1	1
Maternal cousin	1	5	6
Unrelated	0	2	2
Key person doesn't have RA			
All persons in the 6 clusters	1	2	3
Total	17	49	66

DISCUSSION

At this point, it is appropriate to ask about possible errors in the construction of the artificial population in Table 3. First, the population estimate for the proportion of persons truly having rheumatoid arthritis might be faulty for it is a composite from 2 sources. Let us suppose that it varies between 5 and 10 per cent. Then, as can easily be calculated, the specificity would vary between 95 and 99 per cent. Similar estimates of the effect of errors in assessing the frequency of positives on the RA measure can be made, and again within reasonable limits the specificity is not strikingly altered. Further, it can be seen by experimenting with the figures that if the sensitivity has been over-estimated, so also has been the specificity; but again, the effects are small. Between 75 and 100 per cent sensitivity, the specificity ranges between 97 and 99 per cent. There are 2 forces at work that might have led to distortion of the sensitivity. First, it might be under-estimated because it is determined on a clinic population. In a general population sample there would have been relatively far fewer cases needing exclusion; therefore, a smaller proportion of false

negatives, i.e. a greater sensitivity. Secondly, the sensitivity might be over estimated because the clinic sample is loaded with severe classical cases as opposed to borderline cases. We assume that these two forces might have balanced each other out and conclude that this artificial population may, in fact, be a reasonable estimate of the true situation.

Since this RA measure is entirely based on interview information, it seems possible that a tendency to complain might contribute to misclassification. As a first step in testing this hypothesis, a measure of complaint level was constructed from the mean score on a 5-point scale of the stated frequency with which the following symptoms occurred: cough, headache, running nose, pain in the chest, sore eyes, poor appetite, poor digestion, shortness of breath, heart beating fast, dizziness, diarrhea, trouble getting to sleep, trouble staying asleep. Those with the RA measure positive complain of these things somewhat more frequently than those with the measure negative, that the difference is not statistically significant. This led us to take the 21 persons in the first group of the validation sample on whom the complaint data were available and divide them into 2 groups, 9 persons whose interview rating appeared high with respect to their clinic ratings and 12 persons who appeared approximately correctly or slightly under-rated with respect to the clinic rating. The mean complaint level was higher for those persons who were over classified, but the difference was trivial and insignificant ($t=0.2$). This suggests that only a very slight proportion of the association between the RA measure and complaint level could be misclassification of complainers. However, the sample is small and the estimate is, therefore, unstable; so we cannot completely rule out an appreciable classification bias. For this reason the very few variables associated with complaints will receive cautious interpretation when they are related to the RA measure. As a further precaution we have looked for differences between the National sample and the Clinic sample for each variable studied. Where no differences are reported the reader can assume that none exist. Since all the key persons with rheumatoid arthritis in the Clinic sample were clinic diagnosed cases and since these form the bulk of the RA's in this sample the possibility of bias in the RA measure was always entertained when significant differences between Clinic RA's and National sample RA's were found. The first such difference is that the Clinic RA's had more complaints than the National sample RA'S. This, however, does not concern us because clearly the clinic cases are more severe and therefore have more to complain about as a result of their systemic disease.

Finally, we must examine the circumstances under which this measure can be used and under which it cannot. In the first place it is not useful for prevalence estimation, because the usual purposes to which prevalence estimates are put, e.g. planning for health services, demand a greater freedom from bias than any such measure can provide. On the other hand, in the estimation of relative prevalence or in the detection of associations between the disease rheumatoid arthritis and factors of interest, it has great usefulness [5, 19]. As long as there is no association between the sensitivity and the specificity of the measure and the factors of interest, a significant association between the RA measure and a factor is just as significant as if it had been determined using a perfect classification method. One situation, that of complaint level, in which other factors might be associated with the errors of classification has been discussed above and found to be trivial. We cannot exclude the possibility that others may

exist in this group or may be found in different population groups. Of course, something is lost by using imprecise classification, namely the power of the test or the ability to detect an association when it exists.

SUMMARY

Our conclusion is that we have a very usable instrument in the 4-point RA measure, and that if our assumptions are correct, the associations to be presented in subsequent papers are just as true as if the respondents had been examined by a physician equipped with X-ray and laboratory facilities. In some senses this is an affirmation of the medical dictum that the history is by all odds the most important part of a medical evaluation.

REFERENCES

1. LINCOLN, T. A. and COBB, S.: The prevalence of mild rheumatoid arthritis in industry. *J. Occup. Med.* **5,** 10, 1963.
2. COBB, S.: A method for the epidemiologic study of remittent disease. *Am. J. Publ. Hlth* **52,** 1119, 1962.
3. ROPES, M. W., BENNETT, G. A., COBB, S., JACOX, R. and JESSAR, R. A.: 1958 revision of diagnostic criteria for rheumatoid arthritis. *Bull. rheum. Dis.* **9,** 175, 1958.
4. SURVEY RESEARCH CENTER: *Interviewer's Manual.* Institute for Social Research, Ann Arbor, 1966.
5. RUBIN, T., ROSENBAUM, J. and COBB, S.: The use of interview data for the detection of associations in field studies. *J. chron. Dis.* **4,** 253, 1956.
6. BEALL, G. and COBB, S.: The frequency distribution of episodes of rheumatoid arthritis as shown by periodic examination. *J. chron. Dis.* **14,** 291, 1961.
7. CORNFIELD, J.: A method of estimating comparative rates from critical data. *J. Nat. Cancer Inst.* **11,** 1269, 1951.
8. DORN, H. F.: Some applications of biometry in the collection and evaluation of medical data. *J. chron. Dis.* **1,** 638, 1955.
9. ENGLE, A. and BURCH, T. A.: *Rheumatoid Arthritis in Adults, United States 1960-1962.* Ser. 11, No. 17. National Center for Health Statistics, U.S. Government Printing Office, 1966.
10. YOUDEN, W. J.: Index for rating diagnostic tests. *Cancer* **3,** 32, 1950.
11. ROGOT, E. and GOLDBERG, I. D.: A proposed index for measuring agreement in test-retest studies, *J. chron. Dis.* **19,** 991, 1966.
12. SCHILLING, R. S. F., HUGHES, J. P. W. and DINGWALL-FORDYCE, I.: Disagreement between observers in an epidemiological study of respiratory disease, *Br. med. J.* **i,** 65, 1955.
13. STANLEY, J.: Evaluation of a single visit cardiovascular examination. Health Statistics, Ser. D No. 7, Washington, D.C., U.S. Government Printing Office, 1961.
14. GORDON, T. and WATERHOUSE, M.: Hypertension and hypertensive heart disease. *J. chron. Dis.* **19,** 1089, 1966.
15. BIRKELO, C. C., CHAMBERLAIN, E. W., PHELPS, P. S., SCHOOLS, P. E., ZACKS, D. and YERUSHALMY, J.: Tuberculosis case finding. *J. Am. med. Ass.* **133,** 359, 1947.
16. KELLGREN, J. H. and LAWRENCE, J. S.: Radiological assessment of osteoarthrosis. *Ann. rheum. Dis.* **16,** 494, 1957.
17. DUNN, J. P.: *Duodenal Ulcer in Executives with Consideration of Methodological Problems of Formulating an Association Index,* Unpublished Doctoral Dissertation, University of Pittsburgh, 1959.
18. DAWES, R. M. and MEEHL, P. E.: Mixed group validation: A method for determining the validity of diagnostic signs without using criterion groups. *Psychol. Bull.* **66,** 63, 1966.
19. MOTE, V. L. and ANDERSON, R. L.: An investigation of the effect of misclassification on the properties of χ^2-tests in the analysis of categorical data. *Biometrika* **52,** 95, 1965.