



Scandinavian Journal of Primary Health Care

ISSN: 0281-3432 (Print) 1502-7724 (Online) Journal homepage: informahealthcare.com/journals/ipri20

Inter-observer Variation on Registration of Signals of Cancer

Knut Arne Holtedahl

To cite this article: Knut Arne Holtedahl (1987) Inter-observer Variation on Registration of Signals of Cancer, Scandinavian Journal of Primary Health Care, 5:3, 133-139, DOI: 10.3109/02813438709013993

To link to this article: https://doi.org/10.3109/02813438709013993

đ		•
	Т	
	Т	
	Т	

Published online: 12 Jul 2009.



Submit your article to this journal





View related articles 🗹



Citing articles: 1 View citing articles 🗹

Inter-observer Variation on Registration of Signals of Cancer

KNUT ARNE HOLTEDAHL

Kvaløysletta helse- og sosialsenter, Kvaløysletta, Norway

Holtedahl KA. Inter-observer variation on registration of signals of cancer. Scand J Prim Health Care 1987; 5:133-9.

Previous studies have dealt with the role of seven warning signals of cancer in public health education and cancer diagnostics in general practice (1-5). Registrations of warning signals from medical records have some inconveniences common to most retrospective record studies. In the present inter-observer variation study it is found that agreement is good for warning signals in cancer patients when the warning signal has a probable connection with the disease. In control patients, and for more incidentally occurring warning signals in cancer patients, agreement is not so good. It seems that reproducible registrations from records are possible when there is a precise definition of what to look for and the data are relevant to the patient's disease. This study stresses the importance of good record keeping in medical practice.

Key words: comparative study, neoplasm, diagnosis, family practice, statistics.

Knut Arne Holtedahl, Kvaløysletta helse- og sosialsenter, 9100 Kvaløysletta, Norway.

The relationship between seven warning signals of cancer (Table I) and cancer disease has been discussed in previous articles (1-5). Warning signals have been registered in several ways. One method is retrospective study of records from general practice and from hospitals. Any discussion based on results from these studies depends on how reliable is the method. The present study has been carried out to test the validity of registrations from medical records concerning warning signals.

Records are often accessible and are commonly used for retrospective studies. Such studies are not influenced by current registration or control. However, results depend on several other factors:

A. How the notes are taken:

- the patient's ability to describe—verbally or not verbally—his or her symptoms,
- the doctor's ability to listen to and communicate with the patient,
- the extent of the doctor's examination,
- the completeness and accuracy of what the doctor writes in the medical record,
- readability of notes.
 - B. How the record is interpreted:
- the interpreter's personal and educational background,

- the nature of the data to be interpreted and registered.
- the awareness and patience of the interpreter.

From other branches of medicine it is well known that inter-observer agreement may vary in different kinds of diagnostic and evaluation-type settings. The kappa method (6, 7) has been used to compare observations in pathology (8–10), endoscopy (11, 12), neurology (13), as well as observations concerning the general condition of hospital patients (14). Observed agreement always includes some agreement occurring by chance. The calculated kappa coefficient measures agreement beyond chance. Inter-observer studies also have been carried out evaluating the clinical examination of lungs (15), hospital care (16), and problem recording in general practice (17).

MATERIAL AND METHODS

Three general practitioners¹ independently studied the same records from a group of cancer patients and a group of control patients. Comparison of all available records from both general practice and

¹ Knut Arne Holtedahl = Observer 1. Leif Rolfsjord = Observer 2. Terese Fors = Observer 3.

134 K. A. Holtedahl

TT 11 T	~	•		•	,		r •		T C.	37
Table I	Nonon	marnina	CIONALS AT	cancer	according to	o Lanas	storeninoe	n mot	K PPTT.	Norman
I abic I.	00000	waining	Signuis VI	Cuncer	accorants in		1101 01111120		A & I C / I J J	

	Key word (abbreviation used in text)
1. Any sore which does not heal	Sore
2. Lumps anywhere in the body, especially in the breasts,	
and even if they are painless	Lump
3. Abnormal bleeding from body orifices	Bleeding
4. Changes in colour or size of warts and moles	Mole
5. Indigestion or change in bowel habits if this is not	
rapidly normalized	Indigestion
6. Hoarseness or coughing without any apparent reason	Cough/hoarseness
7. Weight loss without any apparent reason	Weight loss

hospitals served as a check on the completeness of what had been written down. Most records were unstructured, with no problem list or titling of progress notes (18). The Cancer Registry of Norway furnished lists of all patients living in the municipality of Tromsø registered between October 1, 1981 and March 31, 1983. Eighty-two of these 331 cancer patients (203 women and 128 men) were found in the files of a previous study where general practitioners in Tromsø registered warning signals at consultation in more than 11000 patients (2). Two patients were not included because their records had been lost. The group of cancer patients then consisted of 52 women and 28 men. Matched control patients were found in the same files as the cancer patients. For each cancer patient the person closest in age and of the same sex was selected. Since all the 160 patients had visited a general practitioner, all of them had at least one record from general practice. In addition, 75 cancer patients (48 women and 27 men) and 39 control patients (28 women and 11 men) had from one to three hospital records. Access to all records was kindly permitted by the owners.

Only information concerning the period between April 1, 1981 (six months prior to the start of diagnosis registration) and cancer diagnosis for each individual patient was considered. For each control patient, record studies included the month of diagnosis for the matched cancer patient. Registration periods thus varied from six to 24 months for different pairs of patients. Warning signals occurring prior to diagnosis were noted on A6 cards for each patient. Warning signals per se were registered irrespective of diagnoses.

For cancer patients the following estimations

Scand J Prim Health Care 1987; 5

were made whenever an observer registered a warning signal:

- is it probable or unprobable that this particular warning signal has a relationship to the cancer disease diagnosed later on?

If the answer was "probable" two more estimations were made:

- did the warning signal occur early or late in the course of the disease?
- was this warning signal potentially useful or not useful for the patient?

Definitions of "early" and "useful" were written down and rehearsed before registrations started. Like in (4), "early" was defined as "a symptom occurring at a time when cure or a rather long remission is possible and not too unlikely. The symptom must stem from the primary tumor or from a metastasis which possibly is solitary". "Late" equals "not early". "Useful" was used when a symptom potentially might contribute to get started a treatment leading to cure or important relief or abbreviation of patient sufferings. Other wise "not useful" was registered. For comparison between cancer patients and control patients, all warning signals registered are taken into account.

The observers, one female and two male, work at three different group practices in Tromsø. Patient populations are predominantly urban but with an important minority of rural/fisherman population. The author has 12 years of experience as a primary care physician, the colleagues three to five years. They were selected by the author not at random but because he knew them as dedicated and conscious practitioners who would carry out the task as accurately as possible.

	Age								
	1-19	20–29	30-39	40-49	5059	60–69	70–89	Total	
Females	1	7	13	4	4	11	12	52	
Males	-	-	2	2	2	12	10	28	
Total	1	7	15	6	6	23	22	80	

Table II. Age and sex distribution of patients

Table III. Inter-observer variation and sex distribution of warning signal registrations (F = Females, M = Males)

	Numi						
	Observer 1		Observer 2		Obse	rver 3	
	F	М	F	м	F	М	
Cancer patients							
Warning signal (s) registered	40	22	43	23	39	21	
No warning signal registered	12	6	9	5	13	7	
Control patients							
Warning signal (s) registered	13	8	18	9	11	9	
No warning signal registered	39	20	34	19	41	19	

A SDS machine was employed in handling the data.

Statistical methods used are: Calculation of overall agreement and kappa values. Analysis is based on observer pairs. Standard error is calculated according to Fleiss (6). McNemar's test for matched pairs. Comparisons of percentages in independent populations.

RESULTS

Table II shows age and sex distribution of the two patient groups. The sex distribution of Tromsø patients does not differ significantly from national figures. Nor is this the case for the patient groups $(0.317 \le p \le 0.5)$.

Table III shows inter-observer variation as to how many patients were registered with at least one

Table IV. Inter-observer variation in number of warning signals registered per patient

	Number of patients										
	Observer 1		Observer 2		Observer 3						
Number of warning signals	Cancer patients	Control patients	Cancer patients	Control patients	Cancer patients	Control patients					
0 .	18	59	14	53	20	60					
1	41	19	42	24	34	16					
2	18	-	17	1	20	3					
3	3	1	7	1	6	-					
4	-	1	-	1	-	1					
Total	80	80	80	80	80	80					

Scand J Prim Health Care 1987; 5

136 K. A. Holtedahl

	Overall agreement	Expected agreement	Kappa	Kappa range	Observer pair with lowest/ highest kappa	
Lump						
Cancer patients	0.85	0.60	0.64	0.61-0.68	A/B	
Control patients	0.94	0.87	0.59	0.39-0.88	C/B	
Bleeding						
Cancer patients	0.89	0.55	0.78	0.76-0.80	A/C	
Control patents	0.93	0.83	0.57	0.42-0.74	B/A	
Indigestion						
Cancer patients	0.84	0.62	0.59	0.56-0.61	A/B	
Control patients	0.93	0.81	0.62	0.56-0.68	C/A	
Cough/hoarseness						
Cancer patients	0.97	0.87	0.77	0.61-0.85	A/B	
Control patients	0.98	0.96	0.40	0 -0.60	C/AB	
Weight loss						
Cancer patients	0.97	0.75	0.88	0.80-0.92	C/AB	
Control patients	0.97	0.95	0.51	0.26-1	BC/A	
All seven warning signals						
Cancer patients	0.93	0.73	0.74	0.72-0.75	A/C	
Control patients	0.96	0.90	0.55	0.49-0.60	C/A	

Table V. Inter-observer agreement on the presence or absence of five warning signals, and on all seven warning signals together. All patients

warning signal. Significantly more warning signals were registered in cancer patients than in control patients (p < 0.001 for each of the three observers).

Table IV shows inter-observer variation in the number of warning signals registered per patient. Altogether 360 warning signals were registered by the three observers, ranging from 86 to 97 warning signals per observer in the cancer group and from 26 to 33 in the control group.

The proportion of patients with warning signals was on the average for the three observers 78% of cancer patients and 28% of control patients. In 63% of the cancer patients warning signals had a probable cancer relationship. The patients had their diagnoses registered during an eighteen month period. This period was divided into three six months periods. Patients diagnosed during the last half year had a longer observation period and might be expected to have experienced more warning signals. For control patients there was a non-significant tendency in this direction, but not for cancer patients. For cancer related symptoms the tendency

Scand J Prim Health Care 1987; 5

was the opposite, perhaps due to an over-representation of younger women with symptom-poor cancer of the cervix in the group with the longest observation period.

Inter-observer agreement on the presence or absence of warning signals is shown for each of the five most registered warning signals in Table V. For the two remaining signals "sore" and "mole" there were too few registrations to justify such analysis separately; no doctor had more than two registrations in each patient group for any of these two warning signals. They are, however, included in the figures for all seven warning signals in Tables V-VII. Figures include overall agreement, agreement expected by chance, and kappa. Figures are given as average values for the three observer pairs,² with range values for kappa, calculated for each observer pair. For the group of cancer patients agreement is good and consistent; "indigestion"

² Observer pair A = Observers 1+2. Observer pair B = Observers 1+3. Observer pair C = Observers 2+3.

Inter-observer variation on registration of signals of cancer 137

Relation				· · · · · · · · · · · · · · · · ·	Observer pair
to the	Overall	Expected			with lowest/
cancer	agreement	agreement	Kappa	Kappa range	highest kappa
Lump					
Likely	0.95	0.66	0.87	0.83-0.94	AC/B
Unlikely	0.89	0.88	0.10	-0.04-0.19	B/A
Bleeding					
Likely	0.91	0.63	0.76	0.73-0.80	B/C
Unlikely	0.89	0.83	0.34	0.23-0.46	B/C
Indigestion					
Likely	0.89	0.72	0.62	0.56-0.70	A/B
Unlikely	0.90	0.83	0.41	0.40-0.42	B/A
Cough/hoarseness					
Likely	0.98	0.91	0.73	0.53-0.88	A/C
Unlikely	0.97	0.95	0.54	0.31-1	AC/B
Weight loss					
Likely	0.98	0.75	0.91	0.84-0.96	C/B
Unlikely	0.99	0.99	0	0	-
All seven warning signals					
Likely	0.95	0.79	0.78	0.76-0.81	AC/B
Unlikely	0.95	0.91	0.38	0.34-0.43	B/A
-					

Table VI. Inter-observer agreement on the presence or absence of five warning signals, with and without a probable relationship with the patient's cancer disease, and on all seven warning signals together. All cancer patients

and "lump" having the lowest average kappas of 0.59 and 0.64 respectively. For control patients there is about the same overall agreement, but with lower registration rates for warning signals agreement by chance is enhanced. This, then, is reflected

in lower kappa values and greater differences between observer pairs. In Table V no observer pair had any tendency of constantly scoring the highest or the lowest kappa values.

For cancer patients the same analysis was carried

Table VII. Analysis of kappa differences, based on average values for all three observer pairs

N = number of warning signal pairs for each observer pair

Warning signals in	N	Kappa	<i>p</i> -value	
Cancer patients Control patients	560 560	0.74+0.08 0.55+0.06	<i>p</i> <0.001	
Cancer patients, probably cancer related Cancer patients, probably not cancer related	560 560	0.78+0.08 0.38+0.04	<i>p</i> <0.001	
Female cancer patients Male cancer patients	364 196	0.76+0.10 0.68+0.14	<i>p</i> =0.369	
Cancer patients: "lump", probably cancer related	80	0.87+0.22	0.107	
probably cancer related	80	0.62+0.22	p=0.107	

Scand J Prim Health Care 1987; 5

138 K. A. Holtedahl

· · · · ·	Can	cer patie	nts		Control actions							
ws	Agre	ement a e observ	ull ers					Agreement all				
	WS, CR		WS, not CR					WS				
	F	М	F	М	No WS	Dis	N	F	м	No WS	Dis	N
Sore	_	_	-	_	77	3	80	_	_	77	3	80
Lump	11	4	-	_	47	18	80	2	1	70	7	80
Bleeding	8	7	1	-	46	18	80	3	1	67	9	80-
Mole	-	-	1	-	78	1	80	-	1	78	1	80
Indigestion	5	2	2	1	49	21	80	3	2	66	9	80
Cough/hoarseness	2	_	1	-	72	5	80	_	-	77	3	80
Weight loss	5	4	-		66	5	80	1	-	75	4	80
Total	31	17	5	1	435	71	560	9	5	510	36	560

Table VIII. Agreement based on number of warning signal triplets

WS = warning signal, N = number of patients = number of warning signal triplets, F = females, M = males, Dis = disagreement, complete or partial, (not) CR = probably (not) cancer-related

out separately for warning signals with and without a probable connection with the cancer disease (Table VI). Observer pair B tends to have the highest kappa values for probably cancer related symptoms, but there are several exceptions.

The most important kappa differences between patient groups, based on average values for all three observer pairs, are summed up in Table VII. Observer agreement is significantly higher for cancer patients than for control patients. So is the case for warning signals which are probably cancer related compared to those probably not cancer related. The differences between the sexes based on all seven warning signals are not significant neither for cancer patients nor for control patients. (Separate calculations for female and male patients, corresponding to Tables V-VI, may be requested from the author.) One might suspect from the kappa values that agreement on "lump" is better than agreement on "indigestion", but comparison of kappas gives p=0.107.

Warning signals noted by all three observers are presented in Table VIII. Among cancer patients two men and seven women had no registrations at all. Fifteen male and 28 female control patients had no registrations. Most of the probably cancer related warning signals were estimated by all three observers to have occurred early, and most of them were considered useful for the patients. Variations in these estimations are complex. Observer 3 had

Scand J Prim Health Care 1987; 5

only two registrations of "late" diagnosis and no "not useful" registrations. The two other observers had about one fourth of their estimations in the "late" group with little disagreement between them on this point. Between the two observers there was agreement upon "useful" for 40 warning signals upon "not useful" for five warning signals, while seven warning signals were considered useful by observer 2 and not useful by observer 1.

DISCUSSION

Cancer related warning signals in cancer patients are well reproducible. Warning signals which are not related to the actual cancer disease are less reproducible. So are warning signals of cancer in non-cancer patients. Agreement on more complexanalysis is uncertain. Overall agreement on "early" and "useful" is good for two of the observers. The third observer has made no "not useful" and only two "late" registrations. This makes it difficult to compare with the two others.

The number of probably cancer related warning signals was not increased for patients with the longest registration period. A long registration period does not automatically mean a long time between the first warning signal and diagnosis. At least doctors should reflect upon a warning signal when it appears, and not wait for a second one.

Retrospective registrations from records seem to

give reproducible results when there is a precise definition of what to look for and the data are relevant to the patient's disease. The findings in this and previous studies (1, 4) indicating that three or four out of five cancer patients before diagnosis experience a warning signal related to the disease, seem quite reliable. Increased precision and accuracy is hardly possible with the records available to-day. Better medical records with problem orientation (18) and use of the ICCP classification system in family practice (19) may lay the ground for increased reliability of retrospective registrations.

ACKNOWLEDGEMENT

The study has been supported by Norsk Forening til Kreftens Bekjempelse, Norway.

REFERENCES

- Holtedahl KA. Kreftdiagnostikk i almenpraksis. I. Kan diagnosen stilles tidligere? (Diagnosis of cancer in general practice. I. Can the diagnosis be made earlier?). Tidsskr Nor Lægeforen 1980; 100: 1219-23 (English summary).
- Holtedahl KA. Kreftdiagnostikk i almenpraksis. III. Almenpraktikeres registrering av faresignaler hos pasienter (Diagnosis of cancer in general practice. III. General practitioners' recording of the "Seven warning signals" in their patients). Tidsskr Nor Lægeforen 1984; 104: 799-800 (English summary).
- Holtedahl KA. Kreftdiagnostikk i almenpraksis. IV. Intervjuregistrering av kreftfaresignaler (Diagnosis of cancer in general practice. IV. Interview-based recording of the occurrence of the "Seven warning signals" in patients in general practice and in the population). Tidsskr Nor Lægeforen 1984; 104: 800-3 (English summary).
- 4. Holedahl KA. Kreftdiagnostikk i almenpraksis. V. Pasienter med og uten kreft (Diagnosis of cancer in general practice. V. Patients with and without cancer: who consults a doctor for the "Seven warning signals"?). Tidsskr Nor Lægeforen 1984; 104:804-6 (English summary).
- Holtedahl KA. Kreftdiagnostikk i almenpraksis. VI. Innholdet i kreftfaresignalene (Diagnosis of cancer in general practice. VI. The content of the warning signals. Is enduring pain relevant as a warning signal?). Tidsskr Nor Lægeforen 1984; 104:806-8 (English summary).
- Fleiss JL. The measurement of interrater agreement. In: Statistical methods for rates and proportions. New York, Chichester, Brisbane, Toronto: Wiley, 1981:212-36.

- Gjørup T, Jensen AM. Kappakoefficienten et mål for reproducerbarhed af nominale og ordinale data (The role of the kappa coefficient in evaluating reproducibility of test results). Nord Med 1986; 101:90-4 (English summary).
- Holman CDJ, James IR, Heenan PJ et al. An improved method of analysis of observer variation between pathologists. Histopathology 1982; 6: 581-9.
- Holman CDJ, Matz LR, Finlay-Jones LR et al. Interobserver variation in the histopathological reporting of Hodgkin's disease: an analysis of diagnostic subcomponents using kappa statistics. Histopathology 1983; 7: 399-407.
- Giard RWM, Hermans J, Ruiter DJ, Hoedemaeker PJ. Variations in histopathological evaluation of nonneoplastic colonic mucosal abnormalities; assessment and clinical significance. Histopathology 1985; 9:535-41.
- Gjørup, T, Agner E, Jensen LB, Jensen AM, Møllmann K-M. The endoscopic diagnosis of gastric ulcer. A randomized clinical trial of interobserver variation and bias induced by knowledge of the radiological diagnosis. Scand J Gastroenterol 1985; 20: 554-8.
- 12. Gjørup T, Agner E, Jensen LB, Jensen AM, Møllmann K-M. The endoscopic diagnosis of duodenal ulcer disease. A randomized clinical trial of bias and of interobserver variation. Scand J Gastroenterol 1986; 21: 261-7.
- Braakman R, Avezaat CJJ, Maas AIR, Roel M, Schouten HJA. Inter-observer agreement in the assessment of the motor response of the Glasgow "coma" scale. Clin Neurol Neurosurg 1977; 80: 100-6.
- Gjørup T, Hamberg O, Knudsen J et al. Interobservatør-overensstemmelsen ved vurdering af patienters almentilstand (Inter-observer agreement in assessing the general condition of patients). Ugeskr Læger 1983; 145: 403-5 (English summary).
- Gulsvik A, Vale JR. Observatørbetinget målefeil ved fysikalsk lungeundersøkelse (Observer variation in physical examination of the chest). Tidsskr Nor Lægeforen 1978; 98: 732-4 (English summary).
- 16. Syse A, Uchermann A, Førde OH. Medisinsk utbytte av sykehusopphold. En evaluering av sykehusoppholdet for pasienter innlagt ved en indremedisinsk avdeling ved et tredelt sykehus (Medical benefit from hospital stay). Tidsskr Nor Lægeforen 1983; 103: 1335-9 (English summary).
- Bentsen BG. The accuracy of recording patient problems in family practice. J Med Educ 1976; 51:311-6.
- 18. Weed LL. Medical records that guide and teach. New Engl J Med 1968; 278: 593-9, 652-7.
- Bentsen BG. Et internasjonalt språk for almenmedisin (An international language for family medicine). Tidsskr Nor Lægeforen 1985; 105: 1849-53 (English summary).

Scand J Prim Health Care 1987; 5