European Journal of Epidemiology (2006) 21: 267–277
DOI 10.1007/s10654-006-0017-7

Interviewer variability – quality aspects in a case–control study

Kerstin J. Blomgren[1,2], Anders Sundström[3], Gunnar Steineck[4,5] & Bengt-Erik Wiholm[1,3,†]
[1]*The Department of Laboratory Medicine, Division of Clinical Pharmacology, Karolinska Institutet at Karolinska University Hospital, Huddinge, Stockholm, Sweden;* [2]*Department of Caring and Public Health, Mälardalen University, Eskilstuna, Sweden;* [3]*Medical Products Agency, Uppsala, Sweden;* [4]*Division of Clinical Cancer Epidemiology, Stockholm City Council, Stockholm, Sweden;* [5]*Department of Oncology and Pathology, Karolinska Institutet, Stockholm, Sweden*

Accepted in revised form 8 February 2006

Abstract. Quality assurance and quality control are important for the reliability of case–control studies. Here we describe the procedures used in a previously published study, with emphasis on interviewer variability. To evaluate risk factors for acute pancreatitis, information including previous diagnoses and medication was collected from medical records and by telephone interviews from 462 cases and 1781 controls. Quality assurance procedures included education and training of interviewers and data validity checks. Quality control included a classification test, annual test interviews, expert case validation, and database validation. We found pronounced variations between interviewers. The maximal number of interviews per day varied from 3 to 9. The adjusted average (95% CI) number of diagnoses captured per interview of cases was 4.1 (3.8–4.3) and of controls 3.5 (3.4–3.7) (excluding one deviating interviewer). For drugs, the average (95% CI) number per interview was 3.9 (3.7–4.1) for cases and 3.3 (3.2–3.4) for controls (excluding one deviating interviewer). One of the fourteen interviewers deviated significantly from the others, and more so for controls than for cases. This interviewer's data 'were excluded. Nonetheless, data concerning controls more frequently needed correction and supplementation than for cases. Erroneous coding of diagnoses and medication was also more frequent among controls. Thus, a system for quality control of coding practices is crucial. Variability in interviewers' ability to ascertain information is a possible source of bias in interview-based case–control studies when ''blinding'' cannot be achieved.

Key words: Interviewer bias, Multi-center studies, Quality assurance, Quality control

Introduction

Quality assurance and quality control are crucial for the validity of a study. Such procedures have been formalized and are well described for clinical trials of medicinal products [1–15]. For epidemiological studies, quality assurance and control is important and of continuous interest especially in epidemiological field studies. Correa et al. [16] reviewed published population-based case–control studies with focus on exposure measurement. They found limited documentation concerning standardization of data collection methods, including how to ascertain exposure.

Quality assurance comprises activities that take place prior to data collection; it includes designing the study, writing the study protocol, developing data collection instruments such as questionnaires, and a manual for data collection, as well as educating and training study monitors [16–22].

Quality control pertains to activities that take place during and after data collection and is a follow-up of quality assurance. Quality control identifies errors in the data during and after data collection [16]. Checking participant eligibility, completeness of data, adherence to protocol, and conformity of data collection performance are typical quality control activities, as are preparing monthly statistical reports and giving regular feedback on quality control results.

Recommendations on how to prevent interviewer-related bias by standardized education and training of interviewers, random call-backs of interview subjects, and by observing interviewer performance have been published [23–32]. We designed a nationwide network for case–control studies of drug-induced

[†]Bengt-Erik Wiholm unexpectedly died on 30th July 2005 after an auto accident.He was specialist in internal medicine as well as in clinical pharmacology, he has been a leader in the field of Drug Safety and Pharmacoepidemiology. He started his carrier in pharmacology and epidemiology at Karolinska Institute in Stockholm Sweden and later on as head for the Pharmacoepidemiology Unit at the Swedish Medical Products Agency in Sweden. Since the last four years he has been employed as Senior Director, Epidemiology at Merck Research and Laboratories in US.

morbidity. Results regarding drugs and pancreatitis have been published [33, 34]. Here, we describe and evaluate the quality assurance and control procedures used in that study, with emphasis on variability among the interviewers.

Material and methods

Study design and study base

A population-based, case–control study including four areas in Sweden with 2.2 million inhabitants was performed between January 1, 1995 and May 31, 1998. The study base consisted of individuals who had telephone were between 20 and 85 years of age, and spoke Swedish. The study base comprised approximately 4.7 million person-years.

Subjects

Cases were patients hospitalized for their first attack of acute pancreatitis at the eight participating hospitals. Patients – even including those with unlisted phone numbers – were identified and monitored from four study sites in different geographical regions. A total of 2453 potential cases were screened; of those, 835 were potentially eligible, 529 were interviewed and 462 were finally included. Controls were drawn quarterly as a random sample from a national population register but restricted to the study areas. The case–control network was designed to study different drug-induced diseases simultaneously with one common pool of controls. Therefore we chose an unmatched design and controls were interviewed continuously. A total of 2245 controls were screened and 1781 were included (Table 1). The central and regional research ethics committees approved the study.

Study interviewers (monitors)

All study interviewers belonged to the Swedish decentralized pharmacovigilance system for which the Medical Products Agency is responsible. In addition to their ordinary duties, including evaluation of suspected spontaneous adverse drug reactions submitted to their centres, they collected all clinical information and performed the interviews in this study. Thus, they were aware that we searched for drugs as possible cause for acute pancreatitis. Seven interviewers participated from the start and seven were added during the study to increase the capacity. The mean age (range) of the interviewers was 42.3 (29–57) years. Two were males and 12 females. Some had received education in epidemiology and research methodology, while others had no such education. One interviewer had a background as a pharmacist and the others were nurses with experience from different medical specialties. Three of the interviewers had experience from at least one other epidemiologic study and another three had experience from randomized clinical studies.

Table 1. Numbers and proportions of screened, potentially eligible, interviewed and finally included cases and controls

Screened excluding monitor A	Cases 2453		Controls 2245	
	N	%	*N*	%
Potentially eligible for interview[a]	835		2206	
Reasons for non-participation				
Failure to establish contact	67	8.0	129	5.8
Dementia	64	7.7	7	0.3
Refusal to participate	55	6.6	187	8.5
Administrative reasons*	94	11.3	1	0.0
Too ill for interview	15	1.8	21	1.0
Deceased	11	1.3	2	0.1
Interviewed subjects[a]	529	63.4	1859	84.3
Excluded after interview				
Previous gallstone	27	5.1	60	3.2
Previous pancreatitis	22	4.2	10	0.5
Cancer in GI tract or pancreas	1	0.2	1	0.1
ERCP pancreatitis	1	0.2		
Other diagnoses than pancreatitis	5	1.0		
Hospitalized past month			6	0.3
Administrative reasons**	8	1.5		
Unreliable interview	3	0.6	1	0.1
Included subjects[a]	462	87.3	1781	95.8

*Specifically: Erroneous exclusions (32 cases), missing medical records (22 cases), transferral to other clinic (21 cases), no study monitor available (17 cases, 1 control) and no permission from treating physician (2cases).
**Specifically: 7 cases interviewed >30 days after hospitalization; 1 case erroneously excluded.
[a]Monitor A excluded.

Quality assurance

Development of study protocol and manual

The study protocol and the interview were developed by the principle investigator (B-EW) in close collaboration with an epidemiologist (AS), the study coordinator (KB), and most of the monitors, who all contributed valuable input from their respective fields of competence.

During the pilot phase several amendments to the interview and study procedures were instituted based on experiences gained.

Monitor meetings and telephone conferences

The study interviewers met with the management group every six months. Between these personal meetings, telephone conferences were held.

Computer program and technical handling of data

The interviews were performed on (laptop) computers with a specially developed program that dynamically displayed questions and stored answers immediately into a database. Information on diseases and drug names was entered and saved in two steps: first, as free text during the interview, and then, after the interview was completed, in a coded format according to ICD-9 and WHO Drug Dictionary. These procedures made it possible for the co-ordinator to check the coding practice of the monitors. Both code systems were available as drop down menus in the computer.

Quality control

Data collection

Potential cases were identified by screening of increased laboratory values of serum amylases, review of the medical record, request to participate, a telephone interview and finally inclusion by the experts. Each potential case was given a study number, an ascertainment form was filled out (Figure 1) and medical records were screened by the study interviewers. For potential cases with obvious exclusion criteria (Table 1), the same information, except date of discharge, was recorded. Remaining potential cases were informed about the study and asked to participate. If he or she agreed, a telephone interview was scheduled within 30 days of hospital admission. The controls were identified and interviewed by telephone continuously during the study period. They received an information letter about the study including an invitation to participate. If they accepted, an interview was scheduled and conducted.

All categorized information noted on the ascertainment forms concerning dates and inclusion/exclusion criteria was compared to the free-text narrative extracted from the medical records. (Figure 1). Completeness of the information recorded was also checked. Corrections of incomplete and erroneous entries were requested.

Each month the number of potential cases and controls screened, interviews performed and planned, and the reasons for primary exclusions and for non-participation in the study were collected and distributed to the study centres.

Interview

Before the interview started, subjects were asked to retrieve all current medicine containers

The interview included seven sections:

(1) clinical and diagnostic information,
(2) demographic data, and all previous hospitalizations,
(3) recent and previous infections,
(4) a detailed medical history, including treatments,
(5) knowledge and experiences of adverse drug reactions,
(6) use of alcohol and tobacco,
(7) occupational and social data.

In the fourth section, 12 disease groups were included, listed in order of frequency based on statistics for outpatient diagnoses, with specific diseases as subgroups. For each disease group, we asked "Do you have or have you had?" followed by an enumeration of the specific diseases, e.g., hypertension or diabetes mellitus. If the response was affirmative and if the disease had been diagnosed by a physician, details on, e.g., onset and duration of the disease were recorded. The follow-up question was: "Have you taken any medicines to treat this condition during the past six months? If the response was "yes", the subsequent questions were about drug name and indication, strength of the drug, dosage, duration of treatment and date/time of the latest use (Figure 2). Finally, as a reminder to the subject not to forget additional drugs used for that condition, a repeat question was posed concerning further drug intake related to that disease group. At the end of the section, there was a set of questions regarding drugs used for varying indications such as pain relievers, vitamins, vaccinations, and herbal remedies.

In Section 5, we asked whether the subjects knew that medicines could cause acute pancreatitis and whether they had been told that their present disease might have been caused by one of their medicines.

The section on the use of alcohol and tobacco was very detailed, see Figure 3.

Classification test

In the beginning of the study, a classification test was conducted to test the interviewers' ability to correctly classify potential cases and to correctly fill out ascertainment forms. Questionnaires with 12 conceivable situations were sent to the interviewers. Each situation had three possible solutions, one being optimal. The results were discussed with the interviewers and the optimal answers given afterwards.

Test interviews

All interviewers were tested yearly by conducting an interview with a fictitious subject. The study coordinator and the coordinator of a parallel study participated in the interview, together with the tested interviewer. One coordinator listened in, acting as an observer, and the other played the role of a fictitious case or control. These tests covered the complete interview procedure, but were not stringently documented. Directly after the test interview, the coordinators discussed and summarized their experiences and gave feedback to the

The National Pancreatitis Study in Sweden Id Nr 46……P……..

Date of Ascertainment:……..

Hospital:……. Personal code number :……

Ward:…….. Name:……

Admission Date:……. Gender:…… Age:…….

Discharge Date:…….. Home adress:…….

Physician:…… Zipcode:…….

1'st contact: (mark with a cross) Phone Number:…….

……(Pers.in hospital)…….Letter…….Phone

Preliminary inclusion criteria fulfilled:

S-amylas twice the upper normal limit: Y / N Abdominal pain: Y / N Admitted to dep of surgery: Y / N

Date fallen ill:……..

Date	CT-scan	U-sound	ERCP

Laboratory values: amylase/lipase

Date	Amylase	AST	ALT	GT	ALP	bilirubin

1'st episode of gallstone:……(Y / N) Performed surgery >5 years:……(Y / N)

Exclusions *(mark with 1, resp 2,3)* Increase of;

……Home district …….. . amylase >72h after admission……. Cytostatic/cancer

……Missed medical records…… Previous gallstones …….. AIDS

……ERCP-induced …….. … Malignancy, GI/pancreas …… Previous pancreatitis

……Excluded by interview……..Intercurrent disease: …… (specify the diagnosis)

Reasons for non participation

…….No Phone ……Dementia ……Hospitalised>30 days

…….Language ……Refusal ……Failure to establish contact

…….Deceased ……Nursing home ……No permission of physician

…….Other reasons:……(specify)……. ……Hospitalised within 30 days

Narrative:……(cont.next page)

Date of letter sent:……..

Figure 1. The National Pancreatitis Study in Sweden.

interviewers. The feedback covered: the introduction of the study, performance speed, adherence to the standardized questionnaire, tendency toward "leading questions", and penetration of incomplete answers.

Evaluation of cases

Two experts in gastroenterology evaluated all primarily included cases on six occasions during the study. They were blinded to all drug exposure and the present hospital in case. They received an abstracted copy of the medical record, including laboratory values and performed examinations together with an extraction from the conducted interview regarding alcohol habits and the history of recent and present disorder from the gastrointestinal tract. They also received the summary of symptoms leading to admission obtained from the interview. Firstly, they assessed every case individually and thereafter they

The screen from a Disease question

Heart: Do You or have You had high blood pressure or any disease in the heart or the blood vessels? (A) (Y,N,U)

Diagnoses (text) Onset Last time used Duration Freq Dx. Code

Have You, during the past six months, taken any medications for this? (B) (Y,N,U)

If Yes, which? Medication (C) *Prompt* *Form*

DrugBox directly:

The Drug Screen (*When the drug name has been entered on the previous page, the specification of the said drug is to be entered)*

Medication: (1) Form: (2) Strength: (3)

Code: (4)

Date Last Used: (6) Hr: Min:

Row DLU *Start* *Freq* *Number* *Dose Text (7)*

……
……

Indication/s for use: (8) *Code*

Have You prior to these past six months used this medication regularly?
If Yes, for how long? (9) (# + D/W/M/Y/U)

Data from calendar ----------Display only----------

W1: W2: W3: W4: (10)

M1: M2: M3: M4: M5: M6: Stop Date

Drug Calendar

Id:……Init……Sex:……Age:……..Adm:……Interv:……

Drug:……..Form:……

Month 1 Week 1-4

W1, DD.MM.YY, Day of week
D1 D2 D3 D4 D5 D6 D7 =……...W1

----------,
DD,MM,YY
D8 D9 D10 D11 D12 D13 D14 =…….W2

And so on, …..(the calendar extends over 182 days, 26 weeks) prior to index

Figure 2. The screen from a disease question.

met together with the co-ordinator for going through their assessments, if their assessments differed, the case was discussed and consensus reached. However, rarely such a situation existed.

Validation of the database

Validation of the database was performed after the study was closed. Reports were constructed for control of plausibility and consistency of data within each block in the interview. Incompletely filled out fields regarding e.g. strength and duration of use of medicines could be completed after discussions with the interviewers, and verifiable errors, i.e. where the coding of e.g. diagnoses could be corrected during the time of validation by comparing the free-text fields with the interviewers' coding or interpretation

Example from the alcohol consumption questions.

Do You drink any alcoholic beverages? ……. ……. (Y, N, U)

If No, have You stopped drinking alcoholic beverages during the past six months?……..(Y, N, U)

(For users and ex-users for less than six months:)

Beverage	Freq	n/Day	LU #+Y,M,W,D	n/LU	Since #+Y,M,W,D
Beer	……	……..…	……	…….	……
Wine	……	……	…….	……	…….
Liquor	……	…….	…….	...……	. ..…….

Freq="How often do you drink"? (Beer, Wine, Liquor), The answer shall be given by some of the alternatives:

Daily, 4-6/W, 1-3/W,. +1/M, -1/M, Never, Unknown.

N/days="How many (glasses of beer/wine, drinks) do you drink on these occasion?

LU= When was the latest occasion you did drink?

n/LU= "At the latest occasion you had a drink, how many glasses/drinks did you have?"

Since= "For how long time have you had these drinking habits"

Figure 3. Example from the alcohol consumption questions.

Statistical methods

Unadjusted and adjusted average number of diagnoses and drugs were calculated with 95% confidence intervals; medians of the number of drugs and diagnoses are presented with inter-quartile ranges (IQR), i.e. the 25th and 75th percentile. The 95% confidence intervals of differences between means were calculated using the *t*-test with unequal variances. The statistical analyses were made in the statistical software package Stata for Windows 8.0 (Stata Corporation, College Station, TX, USA).

Results

Interviewer A conducted 219 control interviews over 69 days, with a maximum of nine interviews in one day (Table 2) – yielding an average of 3.2 control interviews per day compared to the others who on average interviewed 1.6 controls per day.

The average number of diagnoses and the average number of drugs adjusted by age, sex, geographical region and – in the case of average number of drugs – number of diagnoses per interview among cases is given in Table 3. The average number of diagnoses and drugs captured per interview was significantly lower for interviewer A: 3.1 (95% CI 2.5–3.8) and 3.0 (2.4–3.6) for diagnoses and drugs respectively, compared to 4.1 (3.8–4.3) and 3.9 (3.7–4.1) for the other interviewers. The differences between adjusted means of interviewer A and the other interviewers were 1.0 (95% C.I. 0.3–1.7) for diagnoses, and 0.9 (0.3–1.5) for drugs. The deviations were even more pronounced for controls than for cases (Table 4). The adjusted average number of diagnoses and drugs per control

Table 2. Average numbers of control interviews per day among the monitors

Monitor	Interviews	Days with interviews	Average of interviews/ day	Max number of interviews in one day
A	219*	69*	3.2*	9*
B	57	24	2.4	6
C	48	27	1.8	5
D	264	180	1.5	5
E	44	31	1.4	5
F	212	116	1.8	7
G	99	50	2.0	7
H	136	96	1.4	4
I	115	59	1.9	6
J	261	151	1.7	6
K	171	130	1.3	3
L	83	64	1.3	4
M	89	53	1.7	6
N	202	162	1.2	3
Total	1781	1143	1.6	

*Excluded.

Table 3. Number of diagnoses and drugs per interviewer among case interviews

		Age		N/Diagnoses				N/Drugs			
Interviewer*	*n*/Int	Mean (95% CI)		Adjusted mean[a] (95% CI)		Median (IQR)		Adjusted mean[b] (95% CI)		Median (IQR)	
A	50	57.9	(54.3; 61.5)	3.1	(2.5; 3.8)	3	1; 4	3.0	(2.4; 3.6)	2	0.2; 3
D	63	56.2	(51.5; 60.9)	3.6	(3.0; 4.1)	3	2; 5	3.3	(2.9; 3.7)	2	1; 4
G	61	52.4	(48.6; 56.3)	4.0	(3.5; 4.6)	4	2; 5	3.7	(3.2; 4.2)	3	1.5; 5
J	65	56.6	(52.8; 60.5)	4.1	(3.5; 4.7)	4	2; 5.5	4.2	(3.6; 4.7)	4	2; 6
K	22	59.7	(53.4; 66.0)	3.9	(3.0; 4.8)	4	3; 5	4.0	(2.9; 5.1)	4	2; 5
L	46	58.1	(53.8; 62.4)	5.7	(4.8; 6.6)	6	3; 7	2.7	(2.0; 3.4)	3	2; 4.25
M	47	53.2	(48.1; 58.3)	4.2	(3.5; 4.8)	4	2; 6	4.4	(3.6; 5.3)	4	2; 6
N	155	57.3	(54.6; 59.9)	3.7	(3.4; 4.1)	3	2; 5	4.4	(4.0; 4.8)	4	2; 6
Total**	459	56.2	(54.7; 57.7)	4.1	(3.8; 4.3)	4	2; 6	3.9	(3.7; 4.1)	3	2; 5

*Three cases by interviewer F not specified.
**Excluding 50 cases by interviewer A.
[a]Means adjusted by age, sex and geographical region.
[b]Means adjusted by number of diagnoses, age, sex and geographical region.

interview was 1.5 (95% CI 1.0–2.0) and 1.6 (1.3; 1.9), respectively, for interviewer A, compared to 3.5 (3.4–3.7) and 3.3 (3.2–3.4) for the other interviewers. The differences between means of interviewer A and the other interviewers were 2.0 (95% C.I. 1.5–2.5) for diagnoses, and 1.7 (1.4–2.0) for drugs. Thus, the ratio between adjusted average number of diagnoses per case and control was 2.1 (3.1/1.5) for interviewer A compared to 1.2 (4.1/3.5) for the others. The ratio between adjusted average number of drugs per case and control was 1.9 (3.0/1.6) for interviewer A compared to 1.2 (3.9/3.3) for the others. The median age (IQR) of interviewer A was, 56.5 (51.2–66.8) and 53 (40.0–64.0) for cases and controls respectively. The corresponding values for the other interviewers were 56.0 (44.2–70.0) and 53.0 (38.0–65.0). The proportion (95% CI) of males for interviewer A was 56% (49.3–62.8) and 44% (30.0–58.7), for the other interviewers 57% (51.6–60.9) and 49% (46.2–50.9), for cases and controls respectively. Finally, as for geographic area, the average (95% CI) number of diagnoses for subjects from the southern region for interviewer A was 3.2 (2.5–3.9) and 1.8 (1.3–2.3) for cases and controls respectively. For the other interviewers, the corresponding averages for subjects from the same region were 4.2 (3.5–4.8) and 3.2 (3.0–3.5), respectively. Interviewer A also differed from the others regarding "control subjects who denied any use of alcohol": 45.7% compared to the overall 12.9%. Interviewer A made an average of 3.2 control interviews per day compared to the others who interviewed 1.6 controls per day. Interviewer A conducted 219 control inter-

Table 4. Number of drugs and diagnoses per interviewer among control interviews

Interviewer	*n*/Int	Age		N/Diagnoses				N/Drugs			
		Mean (95% CI)		Adjusted mean[a] (95% CI)		Median (IQR)		Adjusted mean[b] (95% CI)		Median (IQR)	
A	219	52.8	(50.6; 55.2)	1.5	(1.0; 2.0)	0	(0; 2)	1.6	(1.3; 1.9)	0	(0; 2)
B	57	56.0	(51.9; 60.2)	4.6	(3.8; 5.3)	4	(2; 6)	3.0	(2.4; 3.5)	3	(2; 5)
C	48	54.3	(49.7; 59.0)	3.3	(2.5; 4.1)	3	(1; 5)	3.5	(2.8; 4.2)	3	(1; 5)
D	264	49.9	(48.9; 51.8)	3.1	(2.6; 3.6)	3	(1; 4)	2.6	(2.2; 3.1)	2	(1; 3.8)
E	44	53.3	(48.4; 58.1)	4.5	(3.5; 5.5)	4	(2; 7)	3.8	(2.7; 4.9)	4	(2; 6)
F	212	53.2	(50.9; 55.5)	4.0	(3.5; 4.5)	4	(2; 5)	3.6	(3.1; 4.1)	3	(2; 5)
G	99	54.8	(51.4; 58.2)	3.4	(2.5; 4.3)	4	(2; 5)	3.6	(2.8; 4.4)	3	(2; 5)
H	136	51.6	(48.9; 54.3)	3.5	(2.3; 4.7)	3	(2; 5)	3.2	(2.1; 4.2)	3	(2; 4)
I	115	53.2	(49.9; 56.5)	3.8	(3.1; 4.4)	3	(2; 6)	3.8	(3.2; 4.3)	4	(2; 5)
J	261	50.7	(48.6; 52.9)	3.7	(2.7; 4.7)	3	(1; 4)	3.5	(2.5; 4.5)	3	(1.5; 5)
K	171	54.4	(51.9; 56.9)	3.7	(3.1; 4.3)	3	(2; 5)	3.3	(2.7; 3.8)	2	(1; 4)
L	83	45.6	(41.8; 49.3)	5.0	(4.3; 5.6)	4	(3; 7)	2.8	(2.2; 3.4)	3	(1; 5)
M	89	56.5	(53.2; 59.8)	3.3	(2.7; 4.0)	3	(2; 5)	3.6	(3.0; 4.3)	3	(2; 5)
N	202	48.9	(46.5; 51.2)	3.3	(3.0; 3.5)	3	(1; 4.2)	4.2	(3.9; 4.6)	4	(2; 6)
Total*	1781	51.9	51.1; 52.7	3.5	(3.4; 3.7)	3	2; 5	3.3	(3.2; 3.4)	3	1; 5

*Total: excluding 219 interviews by interviewer A.
[a]Means adjusted by age, sex and geographical region.
[b]Means adjusted by number of diagnoses, age, sex and geographical region.

Table 5. Average number of drugs and diagnoses among control interviews within different regional areas in Sweden

		Adjusted average number of			
	n/Int	Diagnoses*	95% C.I.	Drugs**	95% C.I.
North	242	3.3	(3.0; 3.6)	3.0	(2.7; 3.3)
Central	1236	3.7	(3.6; 3.9)	3.4	(3.3; 3.5)
South	303	3.2	(2.9; 3.4)	3.7	(3.5; 3.9)
Overall***	1781	3.6	(3.4–3.7)	3.4	(3.3; 3.5)

*Means adjusted by age, sex.
**Means adjusted by age, sex and number of diagnoses.
***Interviewer A excluded.

views over 69 days, with a maximum of nine interviews in one day (Table 2). All interviews performed by interviewer A were excluded from the final analysis concerning etiology [33, 34].

Marked geographical differences regarding reported number of drugs and diagnoses were also noted (Table 5). Regarding the average number of diagnoses, the central part of Sweden demonstrated significantly higher number than those in the other parts. Control subjects from the northern part of Sweden reported an average number of drugs significantly lower than from the other regions, and controls from the southern region reported the highest average number of drugs used.

In the classification test, all interviewers made correct classifications for three out of 12 fictitious potential patients, regarding choice of primary exclusion criterion (i.e. medical reasons) when also secondary criteria (i.e. technical/administrative reasons) were obtained for a case (see Figure 2).

Forty-six ascertainment forms, (1.9%) – based on 2454 potential cases – needed completions, 44 (96%) during the first two years of the study and 2 during the final 2 years. We also identified 33 erroneous exclusions, 25 (76%) of which occurred during the first two years and 8 during the remaining period.

Of all 529 interviewed potential cases, (interviewer A excluded) 67 (13%) were excluded after the interview. The experts screened 497 cases and excluded 35 (7%) of them; they also screened the 11 case patients who died before an interview could be conducted. One of these was excluded.

Overall, erroneous coding of drugs and diseases in the main block of the interview was corrected in 7% of all recorded items for cases and in 18% among the controls. Coding of diseases within the main block in

the interview prompted an action (correction or completion) in 3% of all items among cases and 8% among controls. For drugs, the proportion was 4% for cases and 10% for controls. Indications for drug use were also coded in ICD-9. These had to be corrected in 7% and 21% of cases and controls, respectively.

Discussion

When performing quality control of a large case–control study, we found marked variations between interviewers in numbers of diseases and drugs captured per interview, and in the number of interviews per day. One interviewer deviated significantly from the others.

The variability among interviewers could have several explanations. Some had prior experience of conducting interview studies and some had not. The interviewers also had different backgrounds and personal characteristics that could have contributed to the variability. The impact of interviewer background variables has been discussed albeit infrequently [30, 31]. Johannes et al. [31] found that interviewers with a background in marketing achieved a statistically significant lower response rate to questions regarding recall of subjective or personal information, and to questions that required further probing. None of our interviewers had such a marketing background. It is unclear whether socio-demographic characteristics of interviewers affect response rates [27, 28]. Hox et al. [28] evaluated different socio-demographic and psychological characteristics of interviewers on responses and did not find that these characteristics were of importance for measurement errors, but still concluded that several small differences together could result in pronounced errors. James H. Frey and Sabine Mertens Oischi [32], in their book on interview technique, concluded that ''Good interviewing is the result of quality training combined with interviewer's natural abilities''. Edwards et al. [30] pointed to the importance of quality control measurement with direct feedback to the interviewer on a continual basis to maintain standardized data collection performance. We found uniform training of the interviewers to be of importance for the conformity of their performance [Blomgren et al., submitted].

Drug-utilization, as well as morbidity and diagnostic practices, vary geographically in Sweden [35, 36]. We observed that drugs are lesser used in the northern part of Sweden and that control subjects in the southern part of Sweden reported the highest average number of drugs used. This conforms to information on sales of drugs [37], and it has also been described that the use of psychotropic drugs is high in this area [35].

The interviewer who captured the fewest diagnoses and drugs (Interviewer A in Tables 3 and 4) interviewed subjects predominately from the southern region.

Interviewer A had a lower sensitivity for capturing diagnoses, and subsequently, drug intake among controls than among cases, resulting in a higher ratio between the average number of diagnoses and drugs captured per interview with cases and controls than the other interviewers. A differential measurement error constitutes a potentially detrimental source of bias, and this interviewer's information was excluded from the final analysis of risk factors for pancreatitis. In spite of repeated reminders of the importance of treating cases and controls identically, more completions of missed codes and corrections of erroneous codes were needed for controls than for cases. Thus, it is imperative to have a system where the coding practices can be checked to detect and correct such differences.

In our study the interviewers collected all the clinical information as well as performed the telephone interviews. Thus, they were not blinded to the case or control status of the subjects. This could have contributed to a more conscientious coding of information from cases than controls. One way to reduce such a source of differential measurement errors is to blind the interviewers to the subjects' case status. However, such an approach would have required separate individuals for collecting the clinical information and doing the interviews, a set-up that would demand more resources.

A system for continual evaluation of data would probably have helped us discover the variability between interviewers earlier.

We used test interviews with fictitious subjects to evaluate quality aspects such as adherence to the questionnaire, penetration of incomplete answers, tendency to pose ''leading questions'' and performance speed of the interview. One advantage of this approach was that it made it possible to construct the test interview to focus on various aspects of interviewing and coding. Interviewer A did well in the test interviews, except that she skipped some questions in the questionnaire, an omission she explained as being due to technical reasons. However, two of the other monitors initially had problems with their performance speed, which improved by the second test interview. A prolonged interview can affect both the respondent and the interviewer and thereby affect the data quality. The duration of the interview varied between 20 and 45 min, according to estimates given by the monitors.

Three interviewers initially showed insufficient adherence to the questionnaire and insufficient penetration of incomplete answers. This was discussed with them after both test interviews. Three interviewers also displayed a tendency to leading questions. One of them improved in the second test interview.

Since control subjects with unlisted phone numbers were excluded we cannot ignore a potential situation of bias. However, at the time of present study, an estimate of the frequency of customers with unlisted numbers was 8–10%.

Epidemiologic research does not have the same tradition of working with instructions and guidelines as do clinical trials, which follow Good Clinical Practice guidelines [7]. Good Epidemiological Practice (GEP) has been published and recommended for epidemiological research in the drug, device and vaccine fields [38].

The GEP consists of relatively high-levelled recommendations for performing epidemiological studies. However, recently in 2004, revision of the GEP guidelines has been performed by the International Society for Pharmacoepidemiology, resulting in a more detailed and extended document, the Good Pharmacoepidemiology Practice (GPP) guidelines.

Conclusions

Interviewer related bias, regarding variations associated to the individuals performing the interviews are of interest. Interviewer bias is mainly related to variations in responses that can be associated with the individuals performing the interview. It seems unavoidable that differences in behaviour between interviewers as well as within the same interviewer occur. The challenge is to evaluate, account for, and identify ways to reduce differences between interviewers that actually have impact on the responses [27].

Implications

Based on our experiences, we have the following ambitions for future studies:

- Having a designated person responsible for quality assurance and quality control activities, e.g., a coordinator available from the start, who dedicate the time necessary for all quality-related activities.
- Having a detailed protocol and manual from the start including
  - an action plan for conducting quality control of interviewing and data collection performance, e.g., regular test interviews
  - a description of how quality control of completed questionnaires will be performed for control of data validity, e.g., by randomly selected re-interviews,
  - a definition of what will be regarded as a deviation in monitor performance and how performance will be measured, evaluated, and handled.
- Developing a detailed program for training the interviewers in all aspects of how the study will be conducted, including case ascertainment, data abstraction from medical records, presentation of the study to the subjects, interview technique, data collection, and coding with continual re-training during the whole study period.
- Continuous (monthly) validation of the database during the study.
- Testing all activities in a pilot study, including the above quality control activities.
- Scheduling regular monitor meetings to increase personal involvement and responsibility and to *keep the enthusiasm alive*!

Acknowledgments

The Medical Products Agency, Uppsala, Sweden, provided financial support for this study. We thank Lotta Arnesson, Martin Bäckström, Marianne Claesson, Elisabet Ekman, Kerstin Granberg, Magnus Hansson, Britt Olsson, Matty Persson, Elisabet Stjernberg, Gunilla Stjernlöf, Gunilla Thulin and Ann Titus for work with data collection and Drs. Sven Genell and Svante Sjöstedt for evaluating the cases.

References

1. Ottevanger PB, Therasse P, van de Velde C, et al. Quality assurance in clinical trials. Crit Rev Oncol Hematol 2003; 47(3): 213–235Review.
2. Crerand WJ, Lamb J, Rulon V, Karal B, Mardekian J. Buildning data quality into clinical trials. J AHIMA 2002 Nov–Dec;73(10): 44–6, 48–53, 2 quiz 55–6.
3. Directive 2001/20/EC.Official Journal L 311,28/11/2001: 67–128.
4. ICH Harmonised Tripartite Guideline for Good Clinical Practice. J Postgrad Med 2001; Jul–Sept; 47(3):199–203.
5. Collins CM. Ensuring data quality in the pharmaceutical industry. J AHIMA 1998; 69(6): 34–39.
6. Sharples K, Fleming TR, MacMahon S, Moore A, Reid J, Scoggins B. Monitoring clinical trials. NZ Med J 1998; 111(1072): 322–325.
7. ICH Harmonised Tripartite Guideline for Good Clinical Practice. Brookwood Medical Publications Ltd, UK, 1996.
8. Gassman JJ, Owen WW, Kuntz TE, Martin JP, Amoroso WP. Data quality assurance, monitoring and reporting. Control Clin Trials 1995; 16 (2 Suppl): 104S–136S.
9. Neaton JD, Duchene AG, Svendsen KH, Wentworth D. An examination of the efficiency of some quality assurance methods commonly employed in clinical trials. Stat Med 1990; 9: 115–124.
10. Prud'homme GJ, Canner PL, Cutler JA. Quality assurance and monitoring in the hypertension prevention trials. Hypertension Prevention Trials Research Group. Control Clin Trials 1989; 10 (3 Suppl): 84S–94S.
11. Severe JB, Schooler NR, Lee JH, et al. Ensuring data quality in a multicenter clinical trial: remote site data

entry, central coordination and feedback. Psychopharmacol Bull 1989; 25: 488–490.
12. Tillotson JL, Gorder DD, DuChene AG, Grambsch PV, Wenz J. Quality control in the Multiple Risk Factor Intervention Trial Nutrition Modality. Control Clin Trials 1986; 7 (3 Suppl): 66S–90S.
13. Bagniewska A, Black D, Molvig K, Fox C, Ireland C, Smith J, Hulley S. Data quality in a distributed data processing system: the SHEP Pilot Study. Control Clin Trials 1986; 7: 27–37.
14. Marinez YN, McMahan CA, Barnwell GM, Wigodsky HS. Ensuring data quality in medical research through an integrated data management system. Stat Med 1984; 3: 101–111.
15. Karrison T. Data editing in a clinical trials. Control Clin Trials 1981; 2: 15–29.
16. Correa A, Stewart WF, Yeh HC, Santos-Burgoa C. Exposure measurement in case–control studies: reported methods and recommendations. Epidemiol Rev 1994; 16: 18–32.
17. Christiansen DH, Hosking JD, Danneberg AL, Williams OD. Computer-assisted data collection in multicenter epidemiologic research. The Atherosclerosis Risk in Communities Study. Control Clin Trial 1990; 11: 101–115.
18. Whitney CW, Lind BK, Wahl PW. Quality assurance and quality control in longitudinal studies. Epidemiol Rev 1998; 20(1): 71–80.
19. Cook RR. Overview of good epidemiological practice. J Occup Med 1991; 33: 216–220.
20. Lepore P. The 10 essential activities of the quality assurance unit. Qual Assur 1995; 4: 29–33.
21. Bergman LR. Measurement and data quality in longitudinal research. Eur Child Adolesc Psychiatry 1996; 5 (1 Suppl): S28–S32.
22. Baldwin JK, Hoover BK. Quality assurance and quality control in toxicology: how might these concepts apply to epidemiology? J Occup Med 1991; 33: 1247–1249.
23. Bradburn NM, Sudman S, Blir E. Improving interview Method, Questionnaire Design. San Fransisco CA: Jossey-Bass publishers, 1979, 26 pp–63.
24. Hartge P, Cahill JL, West D, et al. Design and methods in a multi-center case–control interview study. Am J Public Health 1984; 74: 52–56.
25. Wingo PA, Ory HW, Layde PM, Lee NC. The Evaluation of the Data Collection Process for a Multicenter, Population-based, Survey Costs. New York NY: John Wiley and Sons, Inc, 1989, 357 pp–406.
26. Groves RM. Survey Errors, Survey Costs. New York NY: John Wiley and sons, Inc, 1989, 357 pp–406.
27. Fowler FJ, Mangione TW. Standardized survey interviewing: minimizing interviewer-related error (Applied social Research Methods, Vol. 18). Newbury Park, CA: Sage Publications, 1990: 24–32, 96–105, 136–44.
28. Hox JJ, Leeuw ED, Kreft IG. The effect of interviewer and respondent characteristics on the quality of survey data: a multilevel model, Biermar PB, Groves RMLyberg LEMeasurement Errors in SurveysJohn and Sons, IncNew York, NY439461.
29. Wynder Ernst L. Investigator bias and interviewer bias: the problem of reporting systematic error in epidemiology. J Clin Epidemiol 1994; 47(8): 825–827.
30. Edwards S, Slattery M, Mori M, et al. Objective system for interviewer performance evaluation for use in epidemiologic studies. Am J Epidemiol 1994; 140(11): 1020–1028.
31. Johannes CB, Crawford SL, McKinlay JB. Interviewer effects in a cohort study. Am J Epidemiol 1997; 147(5): 429–438.
32. James H. Frey & Sabine Mertens Oishi. How to Conduct Interviews by Telephone and in Person. Thousand Oaks London, New Dehli: Sage Publications International Educational and Professional Publisher, 1995Thousand Oaks, London, New Dehli.
33. Blomgren KB, Sundstrom A, Steineck G, Wiholm BE. A Swedish case–control network for studies of drug induced morbidity – acute pancreatitis. Eur J Clin Pharmacol 2002; 58: 275–283.
34. Blomgren KB, Sundstrom A, Steineck G, Wiholm BE. Obesity and treatment of diabetes with glibenclamide may both be risk factors for acute pancreatitis. Diabetes Care 2002; 25: 298–302.
35. Wessling A, Bergman U, Westerholm B. On the difference in psychotropic drug use between the three major urban areas in Sweden. Eur J Clin Pharmacol 1991; 40: 495–500.
36. Henricsson K, Carlsten A, Ranstam J, et al. Utilisation of codein and propoxyphene:geographic and demographic variation in prescribing, prescriber and recipient categories. Eur J Clin Pharmacol 1999; 55: 605–611.
37. Apoteket AB. Läkemedelsstatistik 1997.
38. Andrews EB, Avorn J, Bortnichak EA, et al. ISPE. Guidelines for good epidemiology practice for drug, device and vaccine research in the United States. Pharmacoepidemiol Drug Safety 1996; 5: 333–338.
39. International Society for pharmacoepidemiology. Guidelines for Good pharmacoepidemiology practices (6pp), 2004.

Address for correspondences: Kerstin J. Blomgren, Mälardalen University, Department of Caring and Public Health, Drottninggatan 16A, S-631 05, Eskilstuna, Sweden
Phone: +46-16-153426; Fax: +46-16-153740
E-mail: kerstin.jorsater-blomgren@mdh.se