Short and long term mortality associated with foodborne bacterial gastrointestinal infections: registry based study

Morten Helms, Pernille Vastrup, Peter Gerner-Smidt, Kåre Mølbak

Abstract

Objectives To determine the excess mortality associated with infections with *Salmonella*, *Campylobacter*, *Yersinia enterocolitica,* and *Shigella* and to examine the effect of pre-existing illness.
Design Registry based, matched cohort study.
Setting Denmark.
Participants 48 857 people with gastrointestinal infections plus 487 138 controls from the general population.
Main outcome measure One year mortality among patients with gastrointestinal infections compared with controls after adjustment for comorbidity.
Results 1071 (2.2%) people with gastrointestinal infections died within one year after infection compared with 3636 (0.7%) controls. The relative mortality within one year was 3.1 times higher in patients than in controls. The relative mortality within 30 days of infection was high in all four bacterial groups. Furthermore, there was excess mortality one to six months after infection with *Yersinia enterocolitica* (relative risk 2.53, 95% confidence interval 1.38 to 4.62) and from six months to one year after infection with *Campylobacter* (1.35, 1.02 to 1.80) and *Salmonella* (1.53, 1.31 to 1.79).
Conclusions Infections with all these bacteria were associated with an increased short term risk of death, even after pre-existing illnesses were taken into account. *Salmonella*, *Campylobacter,* and *Yersinia enterocolitica* infections were also associated with increased long term mortality.

Introduction

Foodborne bacterial infections have a major and perhaps increasing effect on the public health and economy of industrialised countries.[1–4] It is difficult to determine the exact mortality associated with bacterial infections that are usually foodborne. Pathogen specific surveillance systems rarely collect systematic information on outcomes of illness, and outcome specific surveillance systems (such as death certificates) greatly under-report many pathogen specific conditions.[5]

The quantification of the public health impact of bacterial foodborne infections is further complicated by their interaction with chronic underlying diseases and associated conditions.[6 7] We report new estimates of the excess mortality associated with infections with *Salmonella, Campylobacter*, *Yersinia enterocolitica*, and *Shigella* spp. By using data from Danish population based registries, we determined the long term effect on survival adjusted for coexisting illness.

Methods

We obtained data for the study from the national registry of enteric pathogens, the Danish civil registration system, the national registry of patients, and the cancer registry. Bacterial foodborne infections are diagnosed at our institute and 10 local clinical microbiology laboratories. The institute is notified of positive findings and records them in the national registry of enteric pathogens. If a bacterial species or *Salmonella* serotype is found more than once from the same person within six months, only the first positive sample is registered.

We included all patients with culture confirmed infections with non-typhoidal *Salmonella*, *Campylobacter* spp, *Yersinia enterocolitica*, or *Shigella* spp registered between 1 January 1991 and 31 October 1999. To compare the mortality of patients with that of people without known bacterial gastrointestinal infections, we used data from the civil registration system, which assigns a personal identification number to all liveborn children and citizens of Denmark.[8] For every patient, we randomly selected 10 people matched for age, sex, and county of residence who were alive on the date the sample was received. We obtained information on vital status, date of death or emigration, and county of residence for patients and controls. Finally, we obtained data on all hospital discharges, outpatient attendances (since January 1995), and cancer diagnoses up to five years before entry in the study from the national registry of patients and the cancer registry. This allowed us to control for pre-existing illness (comorbidity).

Statistical methods

We created a comorbidity index using the principles described by Charlson et al.[9] This index is a sum of weights corresponding to the number and severity of coexisting illnesses. We first calculated the relative mortality associated with different diagnostic groups, using data from the background population. These relative rates served as weights in the further survival analyses. We then created the index by adding log transformed weights, taking into account multiple discharges before entry into the study. We excluded diag-

Department of Epidemiology Research, Danish Epidemiology Science Centre, Statens Serum Institut, DK-2300 Copenhagen S, Denmark
Morten Helms *research fellow*
Pernille Vastrup *statistician*
Department of Gastrointestinal Infections, Statens Serum Institut
Peter Gerner-Smidt *consultant*
Kåre Mølbak *staff specialist*
Correspondence to: K Mølbak krm@ssi.dk

bmj.com 2003;326:357

Table 1 Comorbidity in 48 857 patients with *Salmonella*, *Campylobacter*, *Shigella,* and *Yersinia* infections, Denmark, 1991-9

	Salmonella (n=26 974)		*Campylobacter* (n=16 180)		*Shigella* (n=1658)		*Yersinia* (n=4045)	
Diagnostic group (weight in survival analysis)	No of cases	Relative risk* (95% CI)	No of cases	Relative risk* (95% CI)	No of cases	Relative risk* (95% CI)	No of cases	Relative risk* (95% CI)
AIDS related illness (3.68)	54	13.16 (8.76 to 19.75)	44	15.17 (9.49 to 24.25)	6	12.01 (3.66 to 39.41)	4	7.98 (2.14 to 29.74)
Metastatic cancers (1.72)	123	2.49 (2.05 to 1.62)	37	2.29 (1.60 to 3.28)	6	4.00 (1.55 to 10.33)	1	0.28 (0.02 to 2.02)
Liver diseases (1.61)	135	3.64 (2.97 to 4.46)	38	1.55 (1.10 to 2.18)	8	3.34 (1.50 to 7.44)	6	1.71 (0.72 to 4.07)
Lymphoma or leukaemia (1.47)	222	3.42 (2.94 to 3.98)	67	2.72 (2.08 to 3.57)	4	2.66 (0.88 to 8.04)	14	3.04 (1.67 to 5.54)
Asthma or COPD (1.34)	633	1.74 (1.60 to 1.90)	315	1.49 (1.32 to 1.68)	15	0.90 (0.53 to 1.53)	82	1.47 (1.16 to 1.85)
Movement disorders† (1.29)	93	1.42 (1.14 to 1.76)	30	0.99 (0.68 to 1.44)	3	0.88 (0.27 to 2.87)	7	1.13 (0.51 to 2.46)
Diabetes (1.21)	639	2.13 (1.95 to 2.32)	227	1.67 (1.45 to 1.92)	14	1.04 (0.60 to 1.81)	57	1.97 (1.48 to 2.62)
Other diseases‡ (1.16)	5 898	1.91 (1.85 to 1.97)	2 790	1.74 (1.67 to 1.82)	212	1.33 (1.14 to 1.55)	761	1.81 (1.66 to 1.97)
No diagnosis (1)	25 246	0.97	15 485	0.99	1604	0.99	3877	0.99

As some patients had more than one disease, the sum of patients with each type of infection is higher than the total number of patients.
*Relative risk of comorbidity in patients with bacterial gastrointestinal infection compared with the general population.
†Parkinson's disease, Huntington's disease, and multiple sclerosis.
‡Cardiovascular diseases (angina, arrhythmia, cerebrovascular, hypertension, myocardial infarction, peripheral vascular); pulmonary diseases other than asthma and chronic obstructive pulmonary disease; endocrine diseases other than diabetes; rheumatological diseases; infections other than HIV, tuberculosis, and gastroenteritis; gastrointestinal other than hepatic and inflammatory bowel diseases.

nostic groups associated with a relative mortality less than 1.2. We forced this index into the survival analyses, so that any difference between the mortality of patients and the general population quantified mortality beyond that attributable to underlying illness.[10–13]

To compare the mortality of patients with that of the general population, we stratified the data so that each stratum contained one patient and 10 controls. We preserved the matching in all analyses by using conditional proportional hazard regression to control for age, sex, and county of residence. The analysis was conducted with SAS software (version 6.12), with proportional hazards regression procedure (PHREG).

Results

During the study, 49 149 patients had bacterial gastrointestinal infections registered, 48 857 (99.4%) of whom could be linked to the civil registry system. Of these patients, 26 974 (55.2%) had *Salmonella* infection, 16 180 (33.1%) *Campylobacter* infection, 4045 (8.3%) *Yersinia* infection, and 1658 (3.4%) *Shigella* infection. A total of 1071 (2.2%) deaths were registered up to one year after infection compared with 3636 (0.7%) deaths among the 487 138 controls. Patients infected with one of the four enteric pathogens had a 3.1 times higher mortality than controls (95% confidence interval 2.89 to 3.33).

A total of 2645 patients had one or more of the diseases included in the comorbidity index. Table 1 shows the number of patients and the various diagnostic groups used in the index, the weights of the diagnostic groups, and the relative risk of belonging to one of the diagnostic groups compared with the reference group. Underlying conditions were more common among patients than in the control group, particularly AIDS related illness, metastatic cancers, and lymphomas or leukaemia. After we adjusted for comorbidity, the relative mortality fell from 3.10 to 2.56 (95 % confidence interval 2.38 to 2.76).

Table 2 summarises the cumulative mortality (Kaplan-Meier estimates) and relative mortality by time since infection. The relative mortality in the 30 days after the episode date ranged from 3.63 to 22.03 for the four bacteria. No excess mortality was seen after 30 days for *Shigella* and 180 days for *Yersinia enterocolitica*, but for *Salmonella* and *Campylobacter*, we found an excess mortality up to one year after infection.

Table 3 shows the relative mortality before and after we adjusted for coexisting illness. After adjusting for comorbidity, we found that mortality in patients infected with *Salmonella dublin* was more than 12 times higher than in the control group. For other *Salmonella* serotypes, *Campylobacter,* and *Yersinia enterocolitica* mortality was 1.86 to 2.88 times higher than in the control group. Infection with *Shigella* species was not

Table 2 Mortality among 48 857 patients with *Salmonella*, *Campylobacter*, *Shigella,* and *Yersinia* infection compared with matched controls from the general population and adjusted for comorbidity, Denmark, 1991-9

	Time after infection (days)			
Type of infection	0-365	0-30	31-180	181-365
Salmonella				
Relative mortality (95% CI)	2.85 (2.61 to 3.10)	13.31 (11.05 to 16.04)	2.22 (1.92 to 2.55)	1.53 (1.31 to 1.79)
Cumulative mortality in patients/controls (%)	3.11/0.97	1.23/0.08	1.08/0.42	0.81/0.47
Campylobacter				
Relative mortality (95% CI)	1.86 (1.56 to 2.20)	4.99 (3.27 to 7.60)	1.85 (1.43 to 2.41)	1.35 (1.02 to 1.80)
Cumulative mortality in patients/controls (%)	1.18/0.52	0.27/0.04	0.50/0.23	0.41/0.25
Shigella				
Relative mortality (95% CI)	1.80 (0.85 to 3.83)	22.03 (4.12 to 117.70)	2.12 (0.58 to 7.80)	0.46 (0.10 to 2.16)
Cumulative mortality in patients/controls (%)	0.66/0.32	0.30/0.01	0.24/0.10	0.12/0.21
Yersinia				
Relative mortality (95% CI)	2.10 (1.40 to 3.16)	3.63 (1.28 to 10.26)	2.53 (1.38 to 4.62)	1.43 (0.74 to 2.79)
Cumulative mortality in patients/controls (%)	0.79/0.38	0.17/0.03	0.35/0.15	0.27/0.19

Table 3 Crude and adjusted relative mortality within one year in 48 857 patients with *Salmonella*, *Campylobacter*, *Yersinia,* and *Shigella* infection compared with control group matched for age, sex, and county of residence, Denmark 1991-9

	No of patients	No (%) of deaths	Relative mortality (95% CI)	
			Crude	Adjusted*
Salmonella (all)	26 974	838 (3.1)	3.44 (3.18 to 3.73)	2.85 (2.61 to 3.10)
S enteritidis	13 967	419 (3.0)	3.34 (2.98 to 3.73)	2.83 (2.51 to 3.18)
S typhimurium	6 988	205 (2.9)	3.63 (3.08 to 4.27)	2.88 (2.42 to 3.44)
S dublin	127	36 (28.3)	17.71 (10.13 to 30.97)	12.35 (6.67 to 22.86)
Other	5 892	178 (3.0)	2.99 (2.52 to 3.55)	2.50 (2.09 to 3.00)
Campylobacter	16 180	190 (1.2)	2.33 (1.98 to 2.73)	1.86 (1.56 to 2.20)
Yersinia	4 045	32 (0.8)	2.16 (1.46 to 3.19)	2.10 (1.40 to 3.16)
Shigella	1 658	11 (0.7)	2.13 (1.11 to 4.08)	1.80 (0.85 to 3.83)

*Adjusted for comorbidity.

associated with higher mortality after we adjusted for comorbidity.

In all, 288 (0.6%) patients were admitted to hospital within 30 days of infection with a diagnosis of an invasive illness (septicaemia, endocarditis, aneurysm, meningitis, pneumonia, abscesses, pancreatitis, or hepatitis). In the control group, 44 (<0.01%) were admitted. The relative mortality among patients with an invasive illness within one year was 17.46 (95% confidence interval 10.11 to 30.17). Among patients with no known invasive illness, the relative mortality was 2.47 times higher than in the control group (2.29 to 2.67).

Of the 48 857 patients with gastrointestinal infection, 46 212 (94.6%) had no other illness included in the comorbidity index. The corresponding figure for the control group was 472 924 (97.1%). Table 4 shows the relative mortality of this group of patients compared with the control group.

Discussion

Most foodborne gastrointestinal infections are self limiting. However, in a subset of patients they can cause severe complications and increased risk of death. Few large systematic studies exist of mortality from foodborne diseases, and the generalisability of the evidence from case reports and outbreaks is questionable. The most recent estimates of mortality were obtained by calculating death rates from the US FoodNet surveillance.[5] The authors assumed that deaths attributable to the foodborne infections were limited to the acute phase of infection and the confounding effect of comorbidity was not taken into account. We were able to examine long term mortality and control for coexisting illness in a large, unselected group of patients. As we used registries created for other purposes, the data should be unbiased.

Effect on mortality

Overall, patients had a 3.10 times higher mortality than the background population within one year of follow up. This figure reflects both acute and long term consequences of foodborne illness as well as the effect of underlying diseases, and it conceals large differences between the bacterial types.

The acute relative mortality was high for all four bacteria after we adjusted for comorbidity (22 for *Shigella,* 13 for *Salmonella*, 5 for *Campylobacter* and 4 for *Yersinia enterocolitica*; table 2). The difference in 30 day cumulative mortality between patients and controls, which correlates to the case fatality rate reported by others, was 1.15% for *Salmonella*, 0.23% for *Campylobacter*, 0.14% for *Yersinia enterocolitica*, and 0.29% for *Shigella*. These figures are in line with the FoodNet case fatality rate for *Yersinia enterocolitica* (0.14%) but are higher than the rates reported for *Salmonella* (0.78%), *Campylobacter* (0.10%) and *Shigella* (0.05%).[5]

We found significant excess long term mortality up to one year after infection with zoonotic *Salmonella* serotypes and *Campylobacter* and up to six months after *Yersinia enterocolitica* infections. By contrast, *Shigella* was mainly associated with death in the acute phase. Unfortunately, valid information on causes of deaths was not available.

Comorbidity

After we adjusted for imbalances in comorbidity, patients infected with *Salmonella, Campylobacter,* or *Yersinia enterocolitica* continued to have a higher mortality than the control group, although the differences were smaller. The high mortality associated with *Salmonella dublin* infections probably reflects its more invasive character.[14–16]

Our comorbidity index is based on discharge diagnoses and on data from outpatient clinics but did not include data from general practitioners. It could be argued that this weakens the index. However, any patient with a pre-existing disease severe enough to alter the outcome of a foodborne infection is likely to have come into contact with a hospital or an outpatient clinic in the five years before infection. Nevertheless, people with other illnesses may have increased

Table 4 Relative mortality within one year in 46 212 patients without any known coexisting disease compared with controls

Infection	No of patients	No (%) of deaths	Relative mortality (95% CI)
Salmonella (all)	25 246	465 (1.8)	2.85 (2.56 to 3.17)
S enteritidis	13 146	234 (1.8)	2.81 (2.42 to 3.26)
S typhimurium	6 518	114 (1.7)	3.01 (2.43 to 3.74)
S dublin	85	15 (17.6)	15.55 (6.57 to 36.80)
Other	5 497	102 (1.9)	2.48 (1.98 to 3.10)
Campylobacter	15 485	115 (0.7)	2.06 (1.68 to 2.53)
Yersinia	3 877	24 (0.6)	2.27 (1.44 to 3.59)
Shigella	1 604	7 (0.4)	1.97 (0.87 to 4.46)

mortality and an independent excess risk of gastrointestinal infection. These people may also be more likely to seek medical attention and have a sample specimen taken than patients without known comorbidity. Among these people, the diagnosis of a gastrointestinal infection may be a marker of excess mortality rather than a contributing cause. However, only a small proportion of patients had a coexistent illness, and the excess mortality was similar in patients with and without underlying illness. Furthermore, there was an excess mortality independent of invasive illness.

Causes of long term mortality

The late excess mortality may have several explanations, including infectious and reactive complications or sequelae, relapses of the initial infection, and reduced efficacy or treatment failure in the case of antimicrobial drug resistance. Complications and sequelae may occur weeks to months after the initial infection and include sequelae of invasive illness (septicaemia, endocarditis, vasculitis, septic arthritis, etc), intestinal perforation, abscesses, and complications of surgery.

The registry did not include multiple diagnoses of the same bacterial species or serotype. We therefore could not examine the importance of relapses. We had only limited data on antimicrobial drug resistance and no information about treatment with antimicrobial drugs and were not able to study this issue. Studies from the United States suggest that treatment with antimicrobials is a risk factor for infection with drug resistant bacteria, and that this interaction may contribute to mortality.[17 18] We have previously shown that quinolone resistance may be associated with excess mortality.[19]

Conclusions

The four foodborne bacterial species we examined were all associated with increased acute mortality. In addition, *Salmonella*, *Campylobacter,* and *Yersinia enterocolitica* were associated with increased long term mortality. Our data suggest that current estimates of the burden of foodborne diseases underestimate the number of deaths from bacterial gastrointestinal infections.

We thank Per Krag Andersen for statistical advice.

Contributors: MH assembled and analysed the data and drafted the article. PV did the statistical analysis and critically revised the article. PG-S was responsible for the registry of enteric pathogens and critically revised the article. KM was responsible for the concept and design of the study, critically revised the article, and is the guarantor. All authors contributed to writing the final manuscript.

Funding: The Research Centre for Environmental Health (Danish Ministry of Health) and the Danish Directorate for Food, Fisheries, and Agro Business (Ministry of Food, Agriculture and Fisheries).

Competing interests: None declared.

1 Todd EC. Epidemiology of foodborne diseases: a worldwide review. *World Health Stat Q* 1997;50:30-50.

What is already known on this topic

Foodborne bacterial infections have a major effect on the public health and economy of industrialised countries

Most estimates of mortality are short term and do not take into account coexisting illnesses

What this study adds

Patients infected with *Salmonella*, *Campylobacter*, *Yersinia*, and *Shigella* had higher 30 day mortality than controls after comorbidity was taken into account

Salmonella, *Campylobacter*, and *Yersinia* infections were also associated with increased long term mortality

The number of deaths from foodborne diseases is likely to be underestimated

2 De Wit MA, Hoogenboom-Verdegaal AM, Goosen ES, Sprenger MJ, Borgdorff MW. A population-based longitudinal study on the incidence and disease burden of gastroenteritis and Campylobacter and Salmonella infection in four regions of the Netherlands. *Eur J Epidemiol* 2000;16:713-8.
3 Potter ME, Tauxe RV. Epidemiology of foodborne diseases: tools and applications. *World Health Stat Q* 1997;50:24-9.
4 Altekruse SF, Cohen ML, Swerdlow DL. Emerging foodborne diseases. *Emerg Infect Dis* 1997;3:285-93.
5 Mead PS, Slutsker L, Dietz V, McCaig LF, Bresee JS, Shapiro C, et al. Food-related illness and death in the United States. *Emerg Infect Dis* 1999;5:607-25.
6 Mauskopf JA, French MT. Estimating the value of avoiding morbidity and mortality from foodborne illnesses. *Risk Anal* 1991;11:619-31.
7 Banatvala N, Cramp A, Jones IR, Feldman RA. Salmonellosis in North Thames (East), UK: associated risk factors. *Epidemiol Infect* 1999;122:201-7.
8 Mosbech J, Jorgensen J, Madsen M, Rostgaard K, Thornberg K, Poulsen TD. [The national patient registry. Evaluation of data quality (in Danish).] *Ugeskr Laeger* 1995;157:3741-5.
9 Charlson ME, Pompei P, Ales KL, MacKenzie CR. A new method of classifying prognostic comorbidity in longitudinal studies: development and validation. *J Chronic Dis* 1987;40:373-83.
10 Elixhauser A, Steiner C, Harris DR, Coffey RM. Comorbidity measures for use with administrative data. *Med Care* 1998;36:8-27.
11 Cleves MA, Sanchez N, Draheim M. Evaluation of two competing methods for calculating Charlson's comorbidity index when analysing short-term mortality using administrative data. *J Clin Epidemiol* 1997;50:903-8.
12 D'Hoore W, Sicotte C, Tilquin C. Risk adjustment in outcome assessment: the Charlson comorbidity index. *Methods Inf Med* 1993;32:382-7.
13 D'Hoore W, Bouckaert A, Tilquin C. Practical considerations on the use of the Charlson comorbidity index with administrative data bases. *J Clin Epidemiol* 1996;49:1429-33.
14 Lester A, Bruun BG, Husum P, Kolmos HJ, Nielsen BB, Scheibel JH, et al. [Salmonella Dublin.] *Ugeskr Laeger* 1995;157:20-4.
15 Lester A, Eriksen NH, Nielsen H, Nielsen PB, Friis-Møller A, Bruun BG, et al. [Bacteremia caused by zoonotic Salmonella types in greater Copenhagen in 1984-1988.] *Ugeskr Laeger* 1990;152:529-32.
16 Fang FC, Fierer J. Human infection with Salmonella Dublin. *Medicine (Baltimore)* 1991;70:198-207.
17 Cohen ML, Tauxe RV. Drug-resistant Salmonella in the United States: an epidemiologic perspective. *Science* 1986;234:964-9.
18 Ryan CA, Nickels MK, Hargrett-Bean NT, Potter ME, Endo T, Mayer L, et al. Massive outbreak of antimicrobial-resistant salmonellosis traced to pasteurized milk. *JAMA* 1987;258:3269-74.
19 Helms M, Vastrup P, Gerner-Smidt P, Mølbak K. Excess mortality associated with antimicrobial drug-resistant, particularly quinolone-resistant, Salmonella typhimurium. *Emerg Infect Dis* 2002;8:490-5.

(Accepted 5 November 2002)

Commentary: matched cohorts can be useful

Stephen Evans

London School of Hygiene and Tropical Medicine, London WC1E 7HT

Stephen Evans *professor of pharmacoepidemiology*

Correspondence to: Stephen.Evans@lshtm.ac.uk

Most *BMJ* readers are familiar with matched case-control studies but fewer will be familiar with matched cohort studies. Case-control studies are based on selecting cases of a disease and then finding people who are as similar as possible to the cases. The study by Helms et al is not a case-control study; people were selected not on the basis of having, or not having, the outcome of interest (in this instance mortality) but on the basis of being exposed or not to something that may affect mortality.

Matched cohort studies have been published in the *BMJ* before—for example, a study examining air bags and deaths of car drivers.[1] Helms et al have used similar methods with Danish national data to look at *Salmonella* (reference 19 of their paper). A common feature of these studies is the existence of a large database in which the individuals who are exposed (to bacterial infection or air bags) can be compared with similar unexposed people. Helms et al used record linkage between databases, obtaining data from microbiology laboratories to define exposed patients and using the national Danish civil registration system to obtain unexposed people from the general population. They also used the registration system to obtain outcome data on subsequent mortality for exposed and unexposed people and two further databases to determine possible confounding from hospital admissions for diseases other than bacterial infection.

The main method of analysis for cohort studies is to use the time taken to an event that is the outcome under study, a survival analysis. The outcome is usually death, but it could be another event such as diagnosis of myocardial infarction or cancer. Cohort studies usually have to be very large to obtain a sufficient number of outcome events. This may make their costs prohibitive, but with electronic databases the costs can be greatly reduced. Similarly, the costs of carrying out matching in cohort studies have restricted their use. Matching prevents the possible association between the matching factors and the exposure at the start of the study, although not necessarily associations occurring as an observational study progresses. Matching should be taken into account in the (conditional) analysis, as has been done by Helms et al.[2]

Matching may not increase statistical power (efficiency) but it does not introduce bias (as it does in case-control studies).[3] With large databases any small loss in efficiency may be unimportant, and the convincing power to the reader of the similarity of the exposed and unexposed cohorts at the start is a gain.

What factors should be used for matching? Helms et al used age, sex, and county of residence. They have used a 1:10 exposed:unexposed ratio. They have also adjusted the survival analysis for comorbidity, based on eight different diagnostic groups. It is possible to match for morbidity or other risk factors, but it would make matching difficult and may not offer any gains. An alternative, used particularly in drug safety, is to match on a "propensity" score.[4] This score measures the likelihood of being given the treatment rather than the likelihood of having the outcome. The purpose is to reduce confounding in either the design or the analysis so that comparisons are valid.

Scandinavia has better national databases than elsewhere, but the United Kingdom has good databases based on general practitioners' computer records. The potential of these is considerable, and matched cohort designs could be used more often. Concerns over confidentiality of records may make this difficult, but it is to be hoped that good epidemiology is not going to be stopped because of misguided ethicists and lawyers.[5]

Competing interests: None declared.

1 Cummings P, McKnight B, Rivara FP, Grossman DC. Association of driver air bags with driver fatality: a matched cohort study. *BMJ* 2002;324:1119-22.
2 Rothman KJ, Greenland S. *Modern epidemiology.* 2nd ed. Philadelphia: Lippincott-Raven, 1998.
3 Greenland S, Morgenstern H. Matching and efficiency in cohort studies. *Am J Epidemiol* 1990;131:151-9.
4 Wang J, Donnan PT. Propensity score methods in drug safety studies: practice, strengths, and limitations. *Pharmacoepidemiol Drug Saf* 2001;10:341-4.
5 Walton J, Doll R, Asscher W, Hurley R, Langman M, Gillon R, et al. Consequences for research if use of anonymised patient data breaches confidentiality. *BMJ* 1999;319:1366.