

AN EMPIRICAL DEMONSTRATION OF BERKSON'S BIAS

ROBIN S. ROBERTS, WALTER O. SPITZER, TERRY DELMORE
and DAVID L. SACKETT

*Department of Clinical Epidemiology and Biostatistics, Faculty of Health Sciences,
McMaster University, Hamilton, Ontario, Canada

**Department of Epidemiology and Health, Faculty of Medicine, McGill University,
Montreal, Quebec, Canada

and supported by National Health Grant 606-21-95, Data generated through Ontario
Ministry of Health Grants, Burlington-DM34, Smithville-DM35, Hamilton-DM53

(Received in revised form May 1977)

Abstract—The importance of Berkson's bias (a systematic bias in the distribution of disease amongst hospitalized patients) in certain types of hospital based research has not been substantiated with empirical data. This paper firstly reviews the theoretical origins of the bias and secondly attempts to provide supportive empirical evidence. The basis of the analysis has been data generated in household surveys designed to capture health utilization information. Sufficient data were available for eight broad clinical conditions and six medications. All possible pairs of associations have been examined in both the entire data set and the hospitalized sub-groups; statistically significant differences in relative risk were then identified. Berkson's theory provides a relatively weak yet statistically significant predictive validity for the change in relative risk from community to hospital.

INTRODUCTION

In 1946, Joseph Berkson, a physician-statistician at the Mayo Clinic, published a celebrated paper which cast doubt on the validity of certain types of epidemiologic research carried out in hospital settings [1]. He suggested that the relative frequency of disease in a group of patients who have entered hospital is inherently biased when compared to the whole population served by the hospital. This phenomenon, attributable to the way in which risks of hospitalization combine in patients who have had more than one disease, has become known as 'Berkson's Fallacy' or 'Berkson's Bias'†. The work is of potentially fundamental importance to the epidemiologic world. Yet, the possible effect of the bias on hospital-based studies that focus on associations of one disease with another (or among clinical entities such as symptoms or injuries) has been largely ignored in the last three decades, presumably due to a lack of supportive evidence for the theory. In the study reported here, an attempt has been made to question this apparent complacency by providing an empirical demonstration of Berkson's Bias.

The idea of a verification arose when one of the co-investigators invoked Berkson's arguments at a conference of the National Institutes of Health. Other epidemiologists rebutted by pointing out, quite rightly, that Berkson had only advanced a theoretical objection, never tested. He then consulted with other members of the team about the feasibility of analyzing clinical data, drug use data and health services utilization information which had been gathered by the McMaster Health Sciences Field Survey Unit in ongoing primary care research. After an encouraging preliminary analysis the data from three studies was brought together to form the basis of the analysis reported here.

A review of Berkson's bias

Berkson centered his discussion on the establishment of an association between the exposure to a suspected causal factor and the presence of a disease. In hospital-based

†Dr. Berkson himself prefers the term 'paradox'.

work, such associations are usually determined through case-control studies. In case-control studies the level of exposure to the putative cause is ascertained in a diseased or *affected* group (the cases) and in a comparative *not affected* group (the controls). The comparative group would be drawn from patients admitted to hospital with a disease which is not thought to have an association with the suspected causal factor under study. For example, patients with accidental bone fractures might be taken as controls when assessing hypertension as a risk factor for skin cancer because there is no plausible association between such fractures and hypertension.

Following Berkson's original model, Fig. 1 shows the three clinical conditions of interest: disease 1, a suspected causal factor thought to be related to the incidence

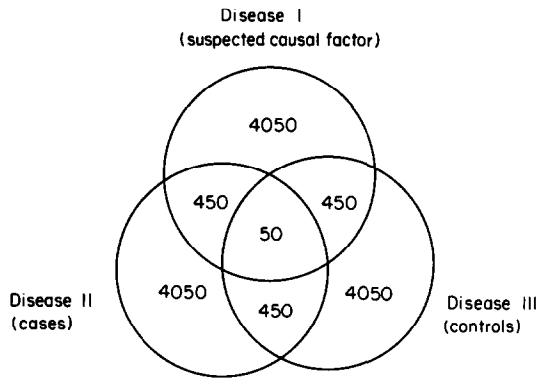


Fig. 1.

of disease 2 which defines cases, and not related to disease 3 which defines controls. In our example above, disease 1 would be hypertension, disease 2 skin cancer, and disease 3 fractures.

The Venn diagram (Fig. 1) depicts the expected distribution of the three diseases as they might arise in a hypothetical free-living population of 50,000 people. It is assumed that the prevalence of each disease is 10% and that they are independent, that is the occurrence of one disease in a person does not affect the likelihood of that person also having one of the other diseases.

A total of 5000 people have disease 1, of whom 450 have disease 1 in combination with disease 2 alone, 450 have disease 1 in combination with disease 3 alone, and 50 people have disease 1 in combination with both disease 2 and disease 3.

An epidemiologist on surveying this population would summarize these data in a 4-fold table as in Table 1 which displays the number of people exposed to the suspected causal factor (disease 1) in both the group of cases having disease 2 and the group of controls having disease 3. In allocating individuals to cells in Table 1, a special decision must be made for subjects exhibiting both disease 2 and disease 3. Because such people

TABLE 1. CASE CONTROL STUDY IN THE GENERAL POPULATION

		Disease 2 (cases)	Disease 3* (controls)
Disease 1 (suspected causal factor)	Present	450 + 50	450
	Absent	4050 + 450	4050

$$\text{Odds ratio} = \frac{(450 + 50) \times 4050}{(4050 + 450) \times 450} = 1.$$

*Without disease 2.

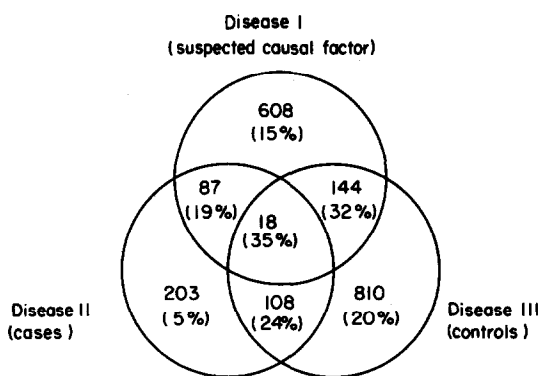


Fig. 2.

are normally included as cases, the precise definition for the control group is 'disease 3 without disease 2'. While relative risk is the traditional statistic used to quantify the association between disease 2 and the suspected causal factor, it cannot be calculated directly from a case control study. However, Cornfield [2] has shown that the odds ratio (ratio of cross products) is a good approximation of relative risk in situations where the prevalences of the conditions of interest are low. In our hypothetical example the estimate by the odds ratio method is $500 \times 4050 / 4500 \times 450$ which is equal to unity. Thus no relationship is apparent between the suspected causal factor (disease 1) and the cases with disease 2. This should not be surprising given the starting assumption of independence.

Taking our example to the next stage, we now consider what happens if the 'selection rates of hospitalization', as Berkson referred to them, are different for the three clinical conditions. Suppose that people with disease 1 are hospitalized during 15% of occasions of illness, people with disease 2 on 5% of occasions, and people with disease 3 on 20% of occasions. If we further assume that these rates follow the laws of probability when two or more diseases are present in a single person, the number of people in the original population who would be expected to enter hospital can be estimated. In Fig. 2 we have shown in parentheses the effective hospitalization rate for each segment of the Venn diagram together with the expected number of hospitalizations rounded to the nearest whole number. For example the 4050 people with disease 2 alone, 5% or 203 would be expected to be hospitalized and thus become cases. The 450 people with both disease 2 and disease 3 would be hospitalized at a combined rate that takes into account the rate for disease 2 alone, 5% and that of disease 3 alone, 20%. The probabilistic union of these two rates produces a hospitalization rate of 24% for these 450 persons, or an expected 108 individuals, who by convention would also be classified as cases. If the search for association were repeated only on those patients from the community who were hospitalized, the data would be as in Table 2. Now the calculated odds ratio of 1.9 indicates an association between cases of disease 2 and disease 1, the suspected causal factor. Since this hypothetical example was set in a community

TABLE 2. CASE CONTROL STUDY IN THE HOSPITAL POPULATION

		Disease 2 (cases)	Disease 3* (controls)
Disease 1 (suspected causal factor)	Present	87 + 18	144
	Absent	203 + 108	810

$$\text{Odds ratio} = \frac{(87 + 18) \times 810}{(203 + 108) \times 144} = 1.9.$$

*Without disease 2.

in which the three diseases were distributed independently from one another, it is clear that the restriction of the search for an association to hospitalized cases and controls introduced a bias. Furthermore, the bias can be attributed to the process of hospitalization alone and particularly to the way in which the differential rates of hospitalization for cases and controls exert their effect together.

The crux of Berkson's argument rests on the assumption that the individual rates of hospitalization for each disease combine together according to the laws of probability. The inevitable result is a different net hospitalization rate for each cell of the 4-fold table so that people with multiple diseases or conditions become over-represented in the hospital population. Berkson did not quantify the effect of the bias in terms of relative risk. He simply discussed the bias in terms of differences in exposure rates and concluded that the bias was most apparent when the three selection rates of hospitalization were markedly different. If the odds ratio estimate of relative risk is used to express the strength of the biased association the algebra simplifies to the point where the expected direction of bias can be described by a simple rule-of-thumb:

'For small control disease prevalence (say below 0.1) the bias decreases relative risk if the case disease hospitalization rate exceeds the control disease hospitalization rate and vice-versa.'

As Berkson pointed out, the bias increases in magnitude with the disparity between hospitalization rates, however only the case and control disease selection rates have any influence. This relationship holds also if the control disease is not a single disease, but 'all other disease' other than the case disease. In this situation the mild dependence on the control disease prevalence drops out of the picture and the bias is solely determined by the case selection rate and the average hospitalization rate for all other diseases.

Such is the theory. We have attempted to confirm or refute it by examining data on the actual experience of communities with morbidity patterns and hospitalization rates measured concurrently.

METHODS

Data sources

Information was available from three separate community based studies of a total of 2784 non-institutionalized adults of age 25 or over. Clinical data, health services utilization data and data about use of drugs were obtained in household surveys of random samples of the persons living in Southern Ontario, over 93% of whom cooperated fully with trained lay interviewers. In one survey, the sampling frame was a census of patients in a suburban family practice; in the other two surveys, the sampling frame was a municipal list of dwellings. Clinical information was elicited the same way in all three surveys: respondents were questioned in their own homes about the presence or absence of health problems in relation to encounters for service and also about clinical events unrelated to service received. Questions about use of drugs prescribed or otherwise, were phrased to enhance recognition of medications frequently used in the survey areas. Responses on health problems were categorized as either a diagnosis or a complaint. Diagnoses were coded using the ICDA-7 system [3] and complaints were coded using the Bain and Spaulding classification [4]. Twenty-seven broad diagnostic (ICDA chapter headings) and complaint groupings were constructed which could be classified as either present or absent together with an indication of whether the respondent had been hospitalized at any time during the 6 months prior to the interview. The full list of 27 clinical entities was reduced to 8 which occurred with sufficient frequency both in the general population and particularly the hospitalized subset (Table 3). The frequency of use of 6 medication groups are also listed at the foot of Table 3. The data were gathered in a way which resembles the manner conventional histories are taken in hospitals particularly with respect to symptoms, illnesses in the recent past, recent use of drugs and recent health care received.

TABLE 3. THE CLINICAL CONDITIONS AND MEDICATION GROUPS

	Total number of patients with the condition	Total number of patients hospitalized (%)
A. Clinical conditions		
1. Allergic, Endocrine System, Metabolic & Nutritional Disease	149	22 (14.7)
2. Mental, Psychoneurotic and Personality Disorders	164	21 (12.8)
3. Diseases of the Circulatory System	193	36 (18.7)
4. Diseases of the Respiratory System	224	20 (8.9)
5. Diseases of the Bones and Organs of Movement	201	23 (11.4)
6. Injuries and Adverse Effects of Chemicals, etc.	101	30 (29.7)
7. Arthritic & Rheumatic Complaints	222	28 (12.6)
8. Fatigue	140	28 (20.0)
B. Medication Groups		
1. A.S.A.	476	51 (10.7)
2. Laxatives	153	20 (13.1)
3. Sleeping pills	60	16 (26.7)
4. Vitamins	359	20 (5.6)
5. Tranquilizers	105	18 (17.1)
6. 'Heart' pills	84	12 (14.3)

Calculations and comparisons

In order to retain all observations in each analysis, we departed from Berkson's style of presentation. Rather than delineating the control group as persons with disease 3, we have delineated the controls simply as those *not affected by disease 2*. Thus in our study, an individual may or may not be affected by the conditions of interest (disease 2). If so, he/she is a case; if not, a control. Concurrently, the individual may or may not be exposed to the suspected causal factor (disease 1). The assignment of conditions to designation as disease 1 or disease 2 has been arbitrary. All possible pairs ($n = 28$) of the eight clinical conditional were analyzed together with all medication/clinical condition pairs ($6 \times 8 = 48$).

The analytic approach was simply to tabulate the data in 4-fold tables, the first analysis including all subjects, the second including only those subjects who had been hospitalized. Since the data came from random population samples we felt it was appropriate to calculate relative risk in the direct manner in addition to the odds ratio estimate. Both statistics are thus included in our analysis. In this way, it was possible to compare the relative risk in the general population with that found in the subset of the same population which was hospitalized. If these two relative risks differed to a degree greater than that which could be expected from sampling variation alone, it could be reasonably attributed to Berkson's bias.

Two approaches have been used to indicate the statistical significance of the observed difference between hospital and general population relative risk. The first method considers the entire study data ($N = 2784$) as a finite population which is stratified according to the presence or absence of the suspected causal factor. The selection of the hospitalized subjects is assumed to occur independently within these groups and thus the distribution of the number of cases within each group follows a hypergeometric distribution. A confidence interval can be placed on the direct estimate of relative risk in the hospitalized subjects using Fieller's theorem [5] and an estimate of its statistical significance obtained. The second method uses the multiple contingency table approach and considers the data as a $2 \times 2 \times 2$ contingency table with factors (1) presence of the suspected causal factor (2) presence of the case disease (3) presence of hospitalization. Using the method due to Bartlett [6], we can test the significance of the second order interaction which, if non-zero, would indicate a difference in odds ratio between hospitalized and non-hospitalized subjects.

Both methods have their advantages and both incorporate distributional approximations. Since the Fieller's theorem approach can be construed as a test of the direct relative risk and the Bartlett approach as a test of odds ratio, we have included both,

to permit the individual reader to select his preferred statistic in this special situation.

A comparison was also made between the magnitude and direction of bias that would occur on the basis of differential hospitalization rates if Berkson's theory were correct and the observed bias in the data.

RESULTS

Our sample included 2784 subjects of whom 257 or 9.2% reported being hospitalized in the 6 months prior to interview. A total of 2564 clinical conditions (diagnoses and complaints) were voiced by these subjects and each was classified into one or more of the 27 groupings. In addition, 60% of subjects reported the use of one or more medications within the 48 hours prior to interview.

The primary analysis relates to the 8 clinical conditions which occurred in 20 or more hospitalized subjects. All possible pairs of conditions were included in the analysis with the designation of suspected causal factor and condition defining cases being arbitrary.

The left hand portion of Table 4 depicts a typical analysis in which we have considered the association between respiratory disease (condition 4) and cases of injury (condition 6). The direct relative risk is 0.98 and the odds ratio estimate also 0.98 indicating no relationship in the general population. Recalculating these statistics in the hospitalized subset provides a direct relative risk of 1.32 and an odds ratio estimate of 1.37 both somewhat higher than the general population. The Fieller's theorem 95% confidence interval for the hospital relative risk is 0.62–1.32 and since it encloses the general population value of 0.98 the observed hospital relative risk is not statistically significant (the actual *P* value for this test is 0.36). Similarly using the multiple contingency table approach the hospitalized and non-hospitalized odds ratios are not significantly different (*P* = 0.57).

This single example is simply one of the 28 possible pairs of conditions which we analyzed. The results of all comparisons appear in Table 5. Here we have displayed direct relative risk and odds ratio in the general population and hospitalized subset together with estimates of statistical significance. Notice that although the contingency table approach is strictly speaking a test of hospitalized versus non-hospitalized odds ratios, to be comparable with the direct relative risks, we have reported here odds ratios for all subjects and the hospitalized subset.

Both direct and odds ratio relative risks show apparently large differences between the entire sample and the hospitalized group. Differences are both positive and negative ranging from +2.23 to -1.82 for the direct relative risk and from +3.00 to -2.09 for the odds ratio. Although agreement between direct relative risk and odds ratio is in general quite good, the odds ratio tends to be consistently larger than the direct relative risk at higher levels of association.

Conventionally statistically significant (*P* < 0.05) differences in hospital direct relative risk were observed in 7 of the 28 comparisons with an additional 2 in the 0.1–0.05 significance range. The contingency table approach is slightly more conservative with

TABLE 4. THE RELATIONSHIP BETWEEN RESPIRATORY DISEASE AND INJURY

	General population				Hospital population				
		injury				injury			
	Yes	No		Yes	No				
Respiratory disease	Yes	8	216	224	Respiratory disease	Yes	3	17	20
	No	93	2467	2560		No	27	210	237
		101	2683	2784			30	227	257

$$\text{Direct relative risk} = \frac{8/224}{93/2560} = 0.98.$$

$$\text{Odds ratio} = \frac{8 \times 2467}{93 \times 216} = 0.98.$$

$$\text{Direct relative risk} = \frac{3/20}{27/237} = 1.32.$$

$$\text{Odds ratio} = \frac{3 \times 210}{27 \times 17} = 1.37.$$

TABLE 5. PAIRWISE RELATIVE-RISK ANALYSIS

Paired clinical conditions	Relative-risk			Odds ratio		P†
	All	Hosp.	P*	All	Hosp	
1-2	1.15	0.53	0.32	1.16	0.51	0.36
1-3	1.49	1.72	0.63	1.55	1.94	0.55
1-4	2.22	2.67	0.71	2.47	3.04	0.72
1-5	1.63	2.97	0.13	1.72	3.55	0.16
1-6	0.92	1.64	0.03	0.92	1.79	0.07
1-7	1.19	1.78	0.32	1.21	1.95	0.36
1-8	1.81	0.40	0.01	1.89	0.37	0.04
2-3	1.95	1.40	0.32	2.09	1.50	0.57
2-4	1.82	0.00	0.04	2.09	0.00	0.01
2-5	1.57	1.69	0.89	1.64	1.80	0.87
2-6	1.18	0.80	0.29	1.20	0.78	0.58
2-7	1.58	1.87	0.68	1.66	2.08	0.67
2-8	2.06	1.34	0.25	2.18	1.41	0.48
3-4	1.46	3.31	0.05	1.52	3.86	0.05
3-5	1.32	1.29	0.96	1.35	1.33	1.00
3-6	1.00	0.68	0.20	1.00	0.65	0.70
3-7	2.02	2.91	0.23	2.20	3.54	0.22
3-8	1.25	2.05	0.06	1.28	2.30	0.07
4-5	1.06	3.29	0.01	1.06	4.06	0.02
4-6	0.98	1.32	0.36	0.98	1.37	0.57
4-7	1.07	1.98	0.14	1.08	2.22	0.22
4-8	2.99	1.42	0.04	3.28	1.50	0.16
5-6	1.90	2.03	0.78	1.96	2.32	0.65
5-7	4.54	4.07	0.72	5.98	5.71	0.92
5-8	0.68	0.78	0.75	0.67	0.76	0.81
6-7	2.06	0.91	0.05	2.26	0.90	0.10
6-8	1.83	1.26	0.24	1.91	1.30	0.74
7-8	2.15	3.27	0.10	2.28	4.18	0.13

*With Fieller's Theorem.

†With Bartlett's approach.

only 4 conventionally significant tables with a further 3 tables in the 0.1-0.05 significance range.

A similar analysis was carried out for all possible medication/clinical condition combinations. Nine of the total of 48 comparisons demonstrated statistically significant differences in direct relative risk between the general population and the hospitalized subset (Table 6). Odds ratio changes showed similar differences although the associated statistical tests were again somewhat more conservative with five attaining significance at better than the 5% level.

Berkson's argument stems from the differential hospitalization rates within each cell of the four fold table. Since our data included both hospitalized numerators and population denominators within each cell, it was possible to calculate these rates for each pair of clinical conditions. Table 7 reports these rates together with the statistical significance of their between cell differences. In only 5 of these comparisons was there no

TABLE 6. MEDICATION/CLINICAL CONDITION COMPARISONS WITH STATISTICALLY SIGNIFICANT CHANGES IN RELATIVE RISK

Medication group	Clinical condition	Direct relative risk			Odds ratio		
		All subjects	Hospitalized	P	All subjects	Hospitalized	P
ASA	Allergy	1.14	0.20	0.02	1.15	0.18	0.02
ASA	Fatigue	1.99	0.76	<0.01	2.09	0.72	0.02
Laxatives	Bones and movement	1.48	4.76	0.04	1.53	5.07	0.06
Laxatives	A & R complaint	1.42	3.40	<0.01	1.48	5.00	0.01
Sleeping pills	Circulatory	4.95	2.56	0.03	6.38	3.27	0.32
Vitamins	Allergy	1.69	0.00	0.02	1.76	0.00	0.01
Vitamins	Injury	0.61	1.79	<0.01	0.61	1.92	0.11
Heart pills	Circulatory	14.77	7.06	<0.01	30.65	19.17	0.47
Heart pills	A & R complaint	2.88	8.82	<0.01	3.46	47.92	<0.01

TABLE 7. HOSPITALIZATION RATES WITHIN THE 4-FOLD TABLE

Clinical condition pair I II	Neither condition cell d	Observed hospitalization rate			Both conditions cell a	Differences in rates P	Predicted rate for cell a			
		Condition 2 alone cell c	Condition 1 alone cell b	Condition 1 alone cell b						
1-2	215/2481	8.7%	20/154	13.0%	21/139	15.1%	1/10	10.0%	0.026	26.1%
1-3	204/2457	8.3%	31/178	17.4%	17/134	12.7%	5/15	33.3%	0.001	27.9%
1-4	219/2436	9.0%	16/199	8.0%	18/124	14.5%	4/25	16.0%	0.111	22.7%
1-5	217/2451	8.9%	18/184	9.8%	17/132	12.9%	5/17	29.4%	0.013	21.4%
1-6	209/2539	8.2%	26/96	27.1%	18/144	12.5%	4/5	80.0%	0.001	36.2%
1-7	211/2427	8.7%	24/208	11.5%	18/135	13.3%	4/14	28.6%	0.011	23.3%
1-8	208/2508	8.3%	27/127	21.3%	21/136	15.4%	1/13	7.7%	0.001	33.4%
2-3	204/2448	8.3%	32/172	18.6%	17/143	11.9%	4/21	19.0%	0.001	28.3%
2-4	216/2419	8.9%	20/201	10.0%	21/141	14.9%	0/23	0.0%	0.044	23.4%
2-5	216/2437	8.9%	20/183	10.9%	18/146	12.3%	3/18	16.7%	0.274	21.9%
2-6	208/2526	8.2%	28/94	29.8%	19/157	12.1%	2/7	28.6%	0.001	38.3%
2-7	212/2418	8.8%	24/202	11.9%	17/144	11.8%	4/20	20.0%	0.101	22.3%
2-8	211/2496	8.5%	25/124	20.2%	18/148	12.2%	3/16	18.8%	0.001	29.9%
3-4	208/2389	8.7%	13/202	6.4%	29/171	17.0%	7/22	31.8%	0.001	22.3%
3-5	202/2408	8.4%	19/183	10.4%	32/175	18.3%	4/18	22.2%	0.001	26.8%
3-6	194/2497	7.8%	27/94	28.7%	33/186	17.7%	3/7	42.9%	0.001	41.3%
3-7	202/2398	8.4%	19/193	9.8%	27/164	16.5%	9/29	31.0%	0.001	24.7%
3-8	200/2462	8.1%	21/128	16.4%	29/181	16.0%	7/12	58.3%	0.001	29.8%
4-5	219/2376	9.2%	18/184	9.8%	15/207	7.2%	5/17	29.4%	0.001	16.3%
4-6	210/2467	8.5%	27/93	29.0%	17/216	7.9%	3/8	37.5%	0.001	34.6%
4-7	213/2357	9.0%	24/203	11.8%	16/205	7.8%	4/19	21.1%	0.145	18.7%
4-8	212/2449	8.7%	25/111	22.5%	17/195	8.7%	3/29	10.3%	0.001	29.2%
5-6	209/2395	8.4%	25/88	28.4%	18/188	9.6%	5/13	38.5%	0.001	35.3%
5-7	214/2419	8.8%	20/164	12.2%	15/143	10.5%	8/58	13.8%	0.200	21.4%
5-8	208/2450	8.5%	26/133	19.5%	21/194	10.8%	2/7	28.6%	0.001	28.2%
6-7	202/2477	8.2%	25/206	12.1%	27/85	31.8%	3/16	18.8%	0.001	40.1%
6-8	203/2552	8.0%	24/131	18.3%	26/92	28.3%	4/9	44.4%	0.001	41.4%
7-8	209/2444	8.6%	20/118	16.9%	20/200	10.0%	8/22	36.4%	0.001	25.2%

r = 0.39

statistically significant variation. As might be expected, the hospitalization rate for the subjects without either of the conditions is in general lowest (cell *d*) and highest for subjects with both diseases (cell *a*). Since Berkson relied upon the probabilistic union to describe the hospitalization rates of patients with 2 or more conditions, it is interesting to verify this assumption in our data. Accordingly in the far right column of Table 7 we have calculated the predicted hospitalization rate for cell *a* based on the union of rates *b* and *c*. The observed and predicted values are not strongly related, but do exhibit a positive correlation of 0.39. Some large discrepancies are present, presumably the result of small sample sizes in cell *a* and the conflicting effects of other influences. In particular, the presence of additional conditions (over and above the 2 of primary interest in each comparison) would severely confuse the rates of hospitalization.

In order to attribute the observed biases specifically to Berkson's phenomenon rather than to general clinical selection bias, we have used Berkson's arguments to predict the expected bias in each of the 28 analyses. This involved firstly, the modification of the theory to reflect the two disease (as opposed to three disease) situation of our comparisons and then the use of the observed hospitalization rates in the unexposed cases and controls (cells *c* and *d*). The rank correlation between the observed bias and expected bias across the 28 comparisons was 0.40. This relatively low value might have been anticipated from the weak validity of the probabilistic union. The rank correlation is, however, positive and statistically significant ($P < 0.025$ one sided).

DISCUSSION

In almost one quarter of the cases in which it was examined (including analysis of drugs as suspected causal factors) a statistically significant distortion of community relative risk was found when the analysis was restricted to hospitalized subjects. Although these results provide substantial support for a Berkson type hospital selection bias, their shortcomings should not be overlooked. Firstly, since 76 individual statistical tests were carried out simultaneously, one would expect at least 3.8 to have been statistically significant by chance alone at the 5% level of significance based on direct relative risk. In our data, 16 discrepancies were observed based on direct relative risk. Furthermore, the 76 analyzed relationships involved only 14 separate diagnostic, complaint, or drug categories and thus are not independent. Although both of these factors temper the real significance of these results, we believe that the differences observed here exceed those that could be attributable to chance.

The differences observed between relative risks in the general population and in the hospitalized sample are unlikely to be entirely due to Berkson's bias. We observed, in fact, what we designate as total bias (B_T). The components of total bias can include: Berkson's bias (B_B) purely the result of a *probabilistic* phenomenon; clinical selection bias (B_{CS}) (which occurs when patients with co-morbidity presentations are more likely to be admitted on *clinical* grounds such as a diabetic on oral hypoglycemics with recent chest pain); other unknown biases (B_U); and random error (e). Therefore, the combined effect can be expressed as

$$B_T = B_B + B_{CS} + B_U + e$$

Our statistical analysis takes e into account. B_U is probably infrequent of lesser magnitude and operates on all groups under assessment. Thus B_B and B_{CS} are the two types of biases contributing most to the observed total bias B_T . If they exert their effect in opposite directions (a plausible situation given the observed reduction of relative risks) the demonstration of bias may be masked. This possibility may explain the relatively low correlation between expected and observed biases.

The apparent limited impact of Berkson's fallacy on epidemiologic research may not simply be due to a lack of empirical evidence. In reviewing the medical and epidemiologic literature for clinical studies in which Berkson's bias may have been an important factor, the number of prime contenders is low. The fact is that few modern studies consider a suspected causal factor which is a disease and thus a force of hospitalization

in its own right. Without the suspected causal factor contributing a direct selective influence one is hard pressed to justify a bias, although one could envisage situations in which the suspected causal factor, while not subject to hospitalization when present alone, could influence the hospitalization decisions if it occurred concurrently with another disease of interest.

Although infrequent in the literature, one *can* find studies which are prone to Berkson's bias. As an example, one might question a case-control study which attempts to link stress with coronary artery disease using organic digestive conditions as a comparative group [7]. In a more recent study [8] benign prostatic hypertrophy was linked to the subsequent development of prostate cancer. Although the strength of the mortality data in this study is impressive, the author suggest that a Berkson type bias may be operating which affects the ensuing detection rate of cancer in patients exhibiting the risk factor. Recent case-control studies exploring a causal association of reserpine and cancer of the breast made it difficult to rule out an association between underlying hypertension and breast cancer [9-11]. Subsequently, a cohort analysis in a very large sample did not demonstrate any effect at all [12] between high blood pressure and the subsequent development of malignancy. Hypertension has also been related to peptic-ulcer [13] and hyperparathyroidism [14]. Diabetes mellitus has been associated with Bell's Palsy [15], gout [16], and idiopathic heart block [17]. In each of these situations there is the potential risk of a Berkson's bias.

Finally, our findings on discrepant risk ratios of drugs as suspected causal factors for certain symptoms or disease should heighten the caution with which hospital-based studies on adverse effects of drug use [18] are interpreted.

Berkson's bias has not received the attention it deserves because its presence is difficult to demonstrate in real situations. To our knowledge, this study is the first attempt at verification and while further confirmation is required, it does provide substantial empirical evidence to support the existence of Berkson type biases. We thus believe that Berkson's bias is an important phenomenon which cannot be overlooked by investigators undertaking epidemiologic or clinical research to show cause-and-effect etiologic relations in hospital populations.

REFERENCES

1. Berkson J: Limitations of the application of 4-fold tables to hospital data. *Biomet Bull* 2: 47-53, 1946
2. Cornfield J: A method of estimating comparative rates from clinical data. Applications to cancer of the lung, breast and cervix. *J Nat Canc Inst* 11: 1269-1275, 1951
3. United States Department of Health, Education, and Welfare: International classification of diseases, adapted for indexing hospital records by diseases and operations. (DHEW Publication No. 719). Government printing office, Washington, DC, Dec. 1962
4. Bain ST, Spaulding WB: The importance of coding presenting symptoms. *Can Med Assoc J* 97: 953-959, 1967
5. Fieller EC: The biological standardization of insulin. *J Roy Stat Soc Suppl* 7: 1-64, 1940
6. Bartlett MS: Contingency table interactions *J Roy Stat Soc Suppl* 2: 248-252, 1935
7. Pearson HES, Joseph J: Stress and occlusive coronary-artery disease. *Lancet* 1: 415-418, 1963
8. Armenian HK, Lilienfeld AM, Diamond EL, Bross ID: Relation between prostatic hyperplasia and cancer of the prostate. *Lancet* 2: 115-117, 1974
9. Boston Collaborative Drug Surveillance Program: Reserpine and breast cancer. *Lancet* 2: 669-671, 1974
10. Armstrong B, Stevens N, Doll R: Retrospective study of the association between use of Rauwolfia derivatives and breast cancer in English women. *Lancet* 2: 672-675, 1974
11. Heinonen OP, Shapiro S, Tuominen L, Tuoumen MI: Reserpine use in relation to breast cancer. *Lancet* 2: 675-677, 1974
12. Garfinkel L, Hammond EC: Breast cancer and hypertension. *Lancet* 2: 1381, 1974
13. Medalie JH, Kahn HA, Neufeld HN *et al*: Association between blood-pressure and peptic-ulcer incidence. *Lancet* 2: 1225-1226, 1970
14. Rosenthal FD, Roy S: Hypertension and hyperparathyroidism. *Br Med J* 4: 396-397, 1972
15. Korczyn AD: Bell's palsy and diabetes mellitus. *Lancet* 1: 108-110, 1971
16. Whitehouse FW, Clearly WJ: Diabetes mellitus in patients with gout. *J Am Med Ass* 197: 113-116, 1966
17. Fairfax AJ, Leatham A: Idiopathic heart block: association with Vitiligo, thyroid disease, pernicious anemia, and diabetes mellitus. *Br Med J* 4: 322-324, 1975
18. Boston Collaborative Drug Surveillance Program. *Lancet* 1: 1399-1404, 1973