International Journal of Epidemiology
© Oxford University Press 1979

Vol 8, No 1
Printed in Great Britain

Cohort Analysis of Lung Cancer in the Netherlands

N M VAN DER HOFF[1]

Van der Hoff N M (Former Medical Registrar, Netherlands Central Bureau of Statistics, Voorburg, Netherlands). Cohort analysis of lung cancer in the Netherlands. *International Journal of Epidemiology* 1979, 8: 41–47.
A cohort analysis of lung cancer mortality in the Netherlands from 1936–1974 is presented. Apart from small corrections, age-specific lung cancer mortality rates can be expressed as the product of a cohort (generation) factor and an age factor. The cohort factors increase until the 1930 birth cohort, after which there is a rapid decline. The age-risk factors increase smoothly with age. Age-specific incidence curves of past and future years will follow the curve of the age-risk factors provided that the cohort-risk factors are unchanging. The method can be applied to other chronic diseases.

INTRODUCTION

Korteweg explained the course of the annual age-specific death rates due to lung cancer in England and Wales, by showing that the increase in the mortality from lung cancer could well be explained by an increase in a generation-bound risk factor (1). Case in his paper on cohort analysis of mortality rates in large populations elaborated on studies dealing with the 'generation effect' made by Springett (2) on tuberculosis, by Korteweg (1) and Haenszel and Shimkin (3) on lung cancer, and concluded that the presentation of mortality rates by year of birth is a better way of providing a basis for biological theories (4).

In the past lack of data has hampered large scale cohort studies. But sufficient data for long enough periods are now available, and by using a computer it has become possible to apply mathematical models. Bjarnason and his colleagues (5) have recently reported the effect of the year of birth on the breast cancer age-incidence curve in Iceland and the mathematical method used (6) has now been applied to mortality data for cancer of the lung in the Netherlands.

MATERIAL AND METHOD

The analysis is based on mortality data for lung cancer in men in the Netherlands, 1936–1974. Women were excluded as numbers of deaths are too small.

The following international classifications of diseases were used.

Period	Revision	Code numbers
1936–1940	4 (1929)	147–155
1941–1949	5 (1938)	142–145
1950–1958	6 (1948)	162–163
1959–1968	7 (1955)	162–163
1969–	8 (1965)	162–1630

All deaths at ages 76 and over were excluded because of unreliability of the diagnosis on death certificates of elderly people, and data for the year 1945 were omitted as numbers of deaths by cause for this particular year are unreliable (7). Tables start from age 20.

The Central Bureau of Statistics analysed the number of deaths due to lung cancer by exact age from 1936. From then onwards one table was made showing the age-specific numbers of death due to lung cancer according to calendar years. This is the basic table from which the cohort table was prepared. The year of birth was calculated by subtracting the age at which death occurred from the year of death. Thus for half the cases the years of birth calculated in this way are not exact, but this was unavoidable. Similarly a table was made for the total male population in the Netherlands according to age and year of birth. These tables are referred to as 1x1 tables and each have 2 145 cells. A further table was constructed by adding numbers per cell of 3 successive ages and 3 successive years of birth. In other words 9 cells of the 1x1 table were compiled in 1 cell. In these new 3x3 tables 1 cell includes 5

[1] Former Medical Registrar, Netherlands Central Bureau of Statistics, Voorburg, Netherlands.
Present address
Peeskesweg 3, Beek Gem., Bergh, Holland.

TABLE 1 *Comparison of the observed and expected numbers of deaths before and after the introduction of the correction factor*

	1 x 1		3 x 3		3 x 3		3 x 3		3 x 3	
	1936 to 1974		1936 to 1974		1947–'49 to 1974		1936 to 1954–'56		1957–'59 to 1974	
	20 to 74 yrs		20–22 yrs to 74–76 yrs		20–22 yrs to 74–76 yrs		20–22 yrs to 74–76 yrs		20–22 yrs to 74–76 yrs	
	without c.f.	with c.f.	without c.f.	with c.f.	without c.f.	with c.f.	without c.f.	with c.f.	without c.f.	with c.f.
chi-square	2148.98	1950.65	311.52	210.72	187.69	137.95	120.74	89.779	132.54	89.176
degrees of freedom	1943	1906	216	203	144	135	90	84	90	84
p in %	< 0.01	23.27	< 0.01	34.46	< 0.01	41.68	1.54	31.56	< 0.01	33.36
(1) total number of cells	2090	2090	266	266	190	190	133	133	133	133
(2) number of cells with $\chi^2 > 3.841$	98	79	19	6	10	5	7	3	9	4
(2) as a percentage of (1)	4.69	3.78	7.14	2.26	5.26	2.63	5.26	2.26	6.77	3.01

The chi-square and the number of cells exceeding the 95 percentile value are the indicators showing degree of agreement between observed and expected values. The year 1945 has been excluded from the 1 x 1 table.
* c.f. – correction factor

successive calendar years: one year 3 times, 2 years twice and one year once*. The year most frequently occurring will be used if reference is made to a period. Subsequent analyses deal with the whole period (1936–1974), the post-war period (1948–74), and in two periods covering 1936–1955 and 1958–1974. The computer department of the Central Bureau of Statistics developed a flexible and comprehensive program so that sizes of tables can easily be altered. All types of tables were tried, the 1x1, 2x2, 3x3, 4x4 and the 5x5 table.

Breslow and Day (6) developed a mathematical method to describe breast cancer incidence in Iceland. A multiplicative model was fitted to a two-dimensional table of age- and cohort-specific rates, in which the age-specific incidence rate for a given cohort was expressed as the product of a cohort factor and an age factor. In the present paper these terms will be referred to as the cohort-risk factor and the age-risk factor. By means of maximum likelihood estimations the method gives estimates of the cohort- and age-risk factors from which a table of expected values can be derived. Comparisons of the observed and expected tables gives goodness of fit tests for the model.

* Copies of the 3x3 table are available from the author or the editorial office.

RESULTS AND DISCUSSION

When the method was applied to lung cancer mortality data in the Netherlands expected and observed values agreed poorly (Table 1, columns without correction factor). The equation $\lambda_{ij} = \theta_i \phi_j$, where λ_{ij} is the incidence rate for the i'th generation at the j'th age group and θ_i and ϕ_j the respective multiplicative contributions, was modified to $\lambda_{ij} = \theta_i \phi_j \epsilon_p$, where ϵ_p is the correction factor for the period p to be added to the equation only after θ_i and ϕ_j have been calculated, as described by Breslow and Day (6). ϵ_p is the factor by which the total number of expected cases per period must be multiplied to make this number equal to the total number of observed cases for the same period. After introducing this correction factor, the agreement between observed and expected number of deaths became reasonably good (Table 1 columns with correction factor).

The cohort-risk factors, the age-risk factors and the correction factors are given in Tables 2, 3 and 4. If, for example, one would like to know the number of deaths due to lung cancer at the age of 50–52 and born in the generation 1920–1922 the corres-

TABLE 2 *Cohort-risk factors*

Year of birth		Year of birth	
1860–'62	0.112951	1908–'10	1.847290
1863–'65	0.155216	1911–'13	1.843390
1866–'68	0.232639	1914–'16	1.943140
1869–'71	0.252886	1917–'19	1.959880
1872–'74	0.276537	1920–'22	1.880310
1875–'77	0.292705	1923–'25	1.882410
1878–'80	0.367278	1926–'28	1.810410
1881–'83	0.421780	1929–'31	2.153950
1884–'86	0.502832	1932–'34	1.897080
1887–'89	0.621961	1935–'37	1.836830
1890–'92	0.771331	1938–'40	1.760390
1893–'95	1.019750	1941–'43	1.431760
1896–'98	1.259360	1944–'46	0.839471
1899–1901	1.396290	1947–'49	0.584584
1902–'04	1.564730	1950–'52	0.000000
1905–'07	1.754270	1953–'55	0.000000

TABLE 3 *Age-risk factors*

Age		Age	
20–22	1.80	50–52	433.00
23–25	3.09	53–55	702.00
26–28	3.24	56–58	1061.00
29–31	5.80	59–61	1520.00
32–34	11.90	62–64	2095.00
34–37	23.60	65–67	2774.00
38–40	43.40	68–70	3517.00
41–43	79.90	71–73	4173.00
44–46	149.00	74–76	5005.00
47–49	255.00		

TABLE 4 *Correction factors*

Period		Period	
1936	1.117190	1957	0.975062
1939	1.040610	1960	0.975010
1942	1.078570	1963	1.017190
1945	0.916114	1966	1.020730
1948	0.986104	1969	1.024360
1951	0.996218	1972	1.010970
1954	0.958057	1974	0.932314

ponding factors in Tables 2, 3 and 4 should be multiplied, giving a rate per million. This rate multiplied by the corresponding population will give the estimated number of deaths: 433 x 1.88031 x 1.01097 x .641406 = 528. This is about 4% less than the number observed (551).

An impression about the agreement between observed and expected rates can be obtained by comparing Figures 1a and 1b. The expected rates per 5 year age group and per period of calendar years are obtained from the analysis of the 1x1

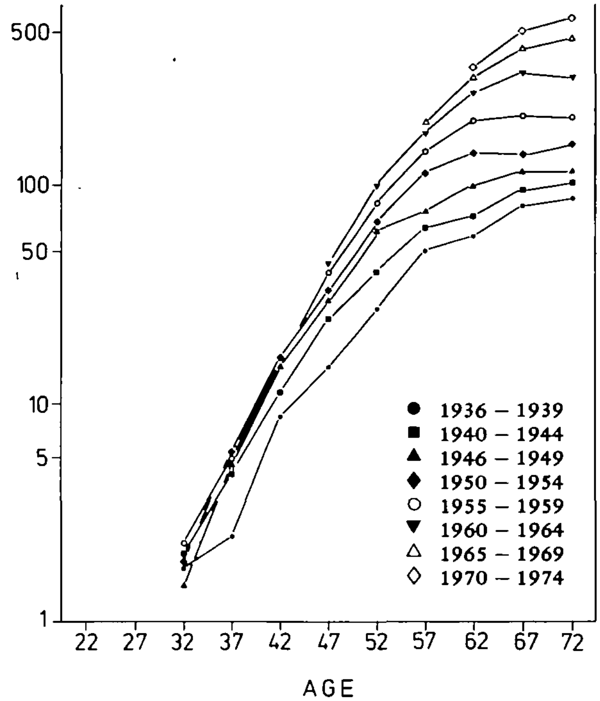

FIGURE 1A *Age-specific mortality rates per 100,000 according to 5-year periods*

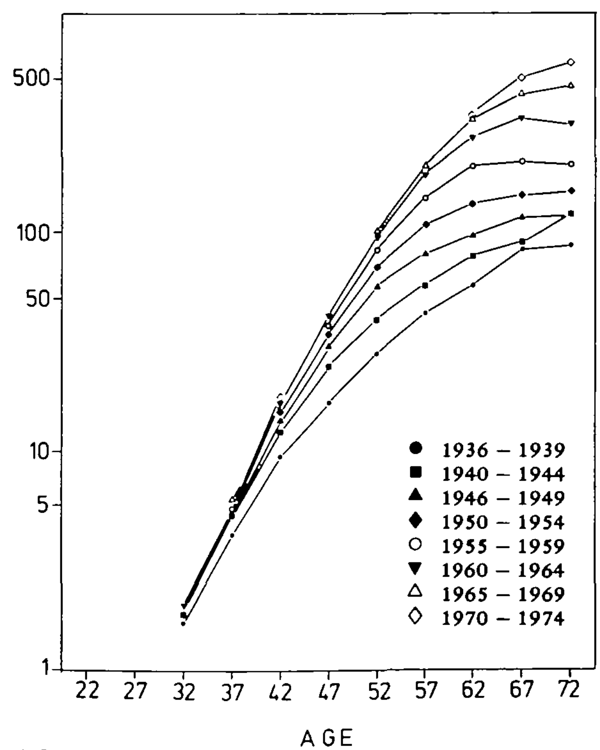

FIGURE 1B *Expected rates, calculated from a 1x1 table and compiled to age-specific mortaility rates according to 5-year periods.*

table adding all the expected numbers of 5 calendar years and age group of 5 years, and dividing this total by the corresponding population.

In Table 1 the sum of the chi square for all cells and the number of cells exceeding the 95 percentile values are given showing the degree of agreement between observed and expected values.

The correction factor

Breslow and Day (6) used formal statistical tools and techniques for choosing an appropriate model for the analysis of the incidence of breast cancer in Iceland. These data on lung cancer are in many respects not comparable with those from Iceland: not because they refer to another form of cancer but rather that they are derived from death certificates on which physicians have written that to the best of their knowledge lung cancer was the underlying cause of death. Information so obtained is obviously different from the carefully scrutinized morbidity data used by Bjarnason and his colleagues (5) and must be recognised as containing a number of errors. It was for this reason that males over 75 years were omitted as also were data for the year 1945 (7). It is not difficult to imagine that the death registration during the whole war period was far less accurate and, finally, it may safely be assumed that during the period 1936–1974, the diagnosis of lung cancer improved considerably.

The chi square goodness-of-fit statistic was used to assess the correctness of the hypothesis. Such statistics, however, should be considered for this type of data as indices to fit rather than as formal tests of goodness-of-fit (11). Initially the goodness-of-fit was rather poor, which was not surprising bearing in mind the possible errors in the data. However, the introduction of the correction factor, which was empirically determined, made the assessment of the goodness-of-fit much better. The chi-square tests should not be considered as significant tests in the usual sense but as tests to support the supposition that the tables of observed and calculated numbers after correction show a reasonable degree of agreement. A better method would have been to challenge the hypothesis and demonstrate that another equation would have given better results. But any other hypothesis would also have had to cope with the imperfectness of the observed numbers. Bearing this in mind it is very difficult to imagine a more practical test.

Although the correction factor refers to certain years or periods, the factors should be considered only as a correction of theoretically calculated rates for each period. But the value of these rates is determined by all observed rates and not only by those observed during a certain period. Hence the correction factor does not play any role in the underlying hypothesis and no absolute value should be attached to it. Moreover, as the age of death at a certain year can only be determined by 2 independent variables, possible influences which one might like to attribute to ϵ_p are already included in θ_i and ϕ_j.

The introduction of the correction factor ϵ_p can be made acceptable only if this factor does not unduly influence λ_{ij}. In other words the less ϵ_p differs from 1, the more likely it becomes that the hypothesis $\lambda_{ij} = \theta_i \phi_j$ is acceptable. Table 2 shows that the most extreme values for the correction factor in the 3x3 table are 1.117 and 0.916 in 1936 and 1945 respectively. The relatively high correction factor for 1936 and the low value for 1974 arise from the boundary problem in cohort analysis. The part-tables show no larger deviations from 1 than the 3x3 table.

The cohort factor

The cohort-factor reflects the complex of influences which from birth contributes to the development at a later age of a certain disease, leading eventually to death. In the study by Bjarnason and his colleagues (5) of breast cancer, and in this study, the hypothesis has been that the incidence of breast cancer and the mortality from lung cancer are determined by such cohort-risk factors together with age-risk factors. In a rapidly changing social, economical and ecological world these complexes of influences might change from generation to generation. Hence, cohort-risk factors might change too and the incidence of breast cancer or mortality from lung cancer accordingly.

The number of years of birth in one cohort should be kept as small as possible, but the results of the analysis of the 1x1 table, especially regarding the cohort-risk factors for younger generations, were too erratic to be useful. The more recent the year of birth, the fewer the number of cells involved and in addition the numbers in some cells of the 1x1 table were very small.

There is a further reason for caution with the younger cohorts. Cancers which to a large extent depend on lasting exposure to certain factors of external origin, may only show a stable cohort-risk factor after the exposure has lasted long enough. The younger the generation the bigger the chance that in the future the cohort-risk factor may change: once the generation involved has been exposed for long enough to the external risk factor,

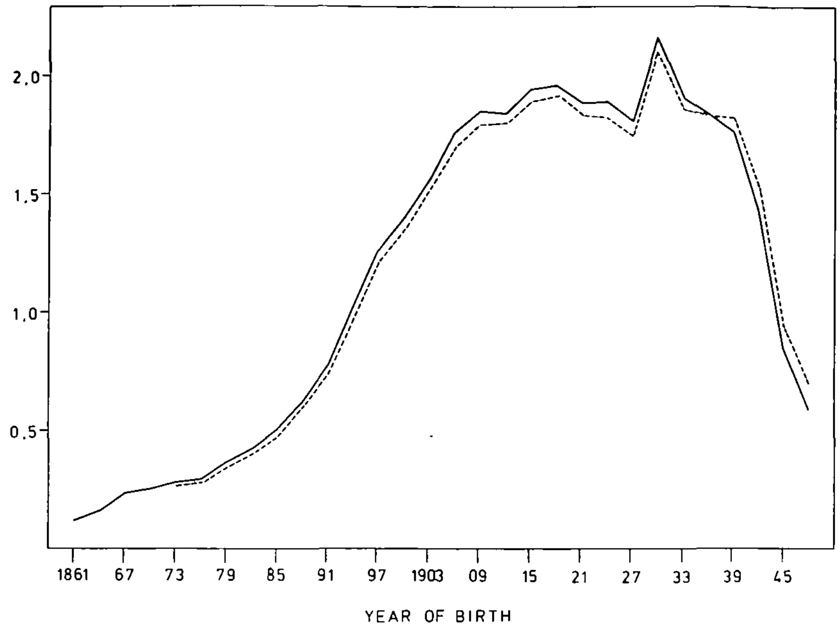

FIGURE 2 *Cobort-risk factors for tbe periods 1936–74 [–] and 1948–74 [.....]*

the cohort-risk factor should be unchanging.

In this study optimal results were obtained from the 3x3 table as shown in Table 2 and Figure 2. To show that these results were not unduly influenced by the bias in the data for the wartime period, an analysis was made of the post-war period only. Figure 2 reveals little difference in the two sets of findings. Both of these show the cohort-risk factors increasing gradually from the oldest generation, reaching the highest point in 1930 and falling sharply thereafter. Bearing in mind what has been said, the possibility that this sharp decline will flatten somewhat in the course of the next few decades must be considered. However, changes to the extent that the trend will be reversed are most unlikely. Thus, the absolute number of deaths due to lung cancer in men will continue to increase in the Netherlands for some years to come as the pre 1930 birth cohorts reach late middle and old age. However, thereafter the number of deaths due to lung cancer should drop sharply.

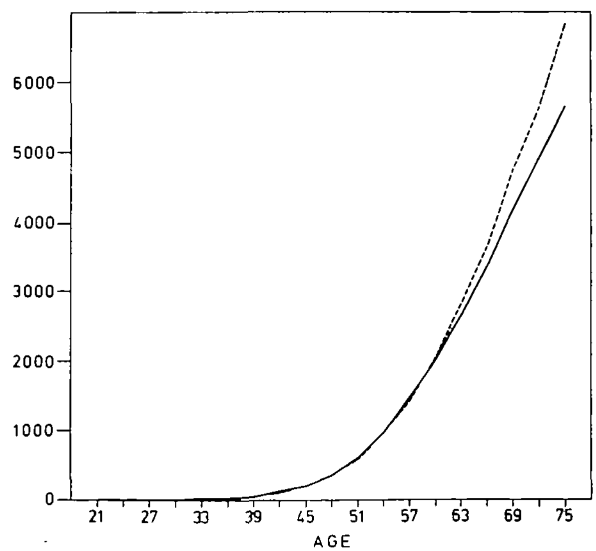

FIGURE 3 *Age-risk factors for tbe periods 1936–55[–] and 1957–74 [.....]*

Tbe age factor

The age-risk factors are shown in Table 3 and Figure 3 for 2 different periods and in Figure 4 for the entire period. This factor represents the relative risk of males dying from lung cancer at any particular age in the Netherlands. It represents the natural development of lung cancer with age, provided no systematic influence alters the course of the curve. That this might occur could be shown by dividing the available data into 2 tables, one covering the period 1936 to 1955 and one for the period 1958 to 1974. The expected rate patterns agree reasonably well with the observed (Table 1). The curves of the age factors are shown in Figure 3. In the second period the age risk factor increases more steeply with age. The most likely explanation for

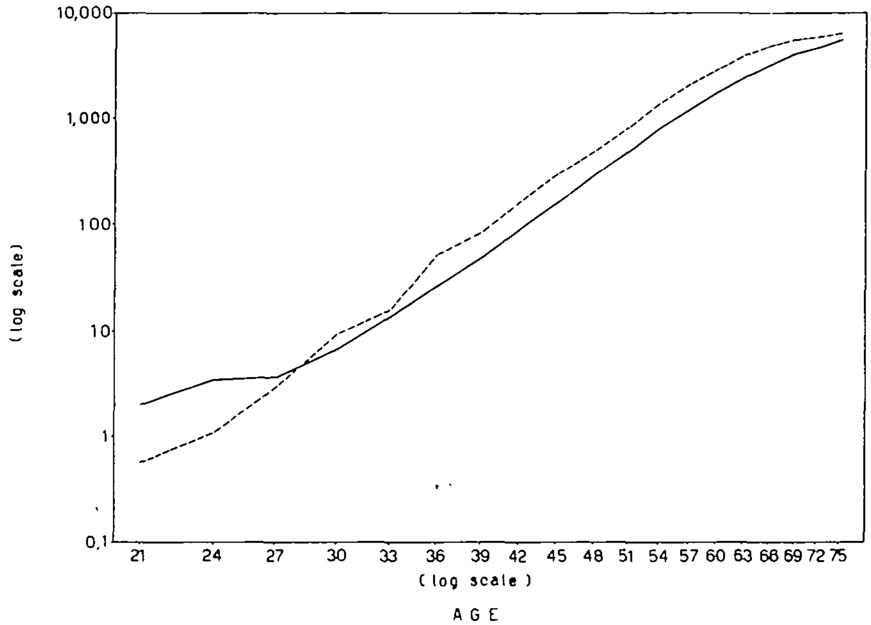

FIGURE 4 *Age-risk factors 1936–74 [—] and age-specific rates per 100,000 1970–74 [.....]*

this is that during the latter period more cases among the older age groups were diagnosed. There is little reason to believe that treatment has influenced the shape of the age-risk curve significantly.

Lung cancer mortality 1970–1974 and its relation to the age-risk factors.

If standardized age-specific mortality rates are determined by the product of two factors, the cohort-risk factor and the age-risk factor, the annual age-specific mortality will follow a course parallel to the curve of the age-risk factors, provided that the cohort-risk factors have remained constant for all the generations involved. If this has not been the case, the age-specific mortality curve for a certain period will deviate from the age-risk factor curve. This can be demonstrated by plotting log age-specific mortality rates and the log age-risk factors against log age together (Figure 4). The fluctuations in the distance between the two curves represent the fluctuations of the log cohort-risk factors of the generations concerned. Because the use of logarithms tends to flatten the real differences, the changes in the cohort-risk factors as shown in Figure 2, appear far less obvious in Figure 4. To examine individual years there is no need to use logarithmic paper. If the ratio age-risk-factor/age-specific incidence is constant there is no reason to believe that the real incidence of the disease is changing. If the ratio is increasing with the years the cohort-risk factor is on the decline and if the ratio is on the decrease the reverse is the case.

From these results it appears that the annual mortality from lung cancer in a given population in the first place depends on the course of the curve of the age-risk factors and secondly on the risk-factors particular to each generation. This analysis for the Netherlands suggests that, starting with the generations born about 1890, exposure to the complex of external influences leading to lung cancer increased. There is of course the possibility that the number of influences increased as well. The generations born after 1930 appear to be less exposed. This method of analysis can be applied to mortality data for other conditions.

ACKNOWLEDGEMENTS

The author is indebted to Dr C A W Citteur and Mr F M Kellenbach of the Netherlands Central Bureau of Statistics and Dr N E Day of the International Agency for Research on Cancer, Lyon, for their helpful advice and assistance with the preparation of this paper and to Mr L Willems and Mr G Degen for their devoted technical assistance.

REFERENCES

(1) Korteweg R. The age curve in lung cancer. *British Journal of Cancer* 5: 21, 1951.
(2) Springett V H. A comparative study of tuberculosis mortality rates. *Journal of Hygiene* 48: 361, 1950.
(3) Haenszel W and Shimkin M B. Smoking patterns and epidemiology of lung cancer in the United States; are they compatible? *Journal of the National Cancer Institute* 16: 1417, 1956.

(4) Case R A M. Cohort analysis of mortality rates as an historical or narrative technique. *British Journal of Preventive and Social Medicine* 10: 150, 1956.
(5) Bjarnason O, Day N, Snaedal G and Tulinius H. The effect of year of birth on the breast cancer age-incidence curve in Iceland. *International Journal of Cancer* 15: 689, 1974.
(6) Breslow N E and Day N E. Indirect standardization and multiplicative models for rates, with reference to the age adjustment of cancer incidence and relative frequency data. *Journal of Chronic Diseases* 28: 1, 1975.
(7) Korteweg R. Mortality from lung cancer in the Netherlands during and after the last war. *British Journal of Cancer* 8: 34, 1954.
(8) Bishop Y M M, Fienberg S and Holland P W. Model selection and assessing closeness of fit: practical aspects, p 311. In Discrete Multivariate Analysis: Theory and Practice. MIT Press, Cambridge, 1975.

(Revised version received 13 July 1978)

WHO Publications

The following publications have been received :

The Management of Nutritional Emergencies in Large Populations. C de Ville de Goyet, J Seaman and U Geijer, Geneva, World Health Organization, 1978, pp 98, S W fr 16.

Field Guide to the Detection and Control of Xerophthalmia. A Sommer, Geneva, World Health Organization, 1978, pp 47, Sw fr 10.

Application of the International Classification of Diseases to Dentistry and Stomatology, 2nd Edition, Geneva World Health Organization, 1978, pp 147, Sw fr 20.

Environmental Health Criteria 6 : Principles and Methods for Evaluating the Toxicity of Chemicals Part 1. Published under the joint sponsorship of the United Nations Environment Programme and the World Health Organization. Geneva, World Health Organization, 1978, pp 272, Sw fr 28.

World Health Statistics Vol 31, No 3, Geneva. World Health Organization, 1978. Special Subject :present trends in mortality in the age group 35–64 in selected developed countries between 1950 and 1973.

World Health Statistics Vol 31, No 4. Geneva WHO, 1978. Special Subject sex differentials in cancer mortality and morbidity.

Directory of On-going Research in Cancer Epidemiology 1978. Eds C S Muir and G Wagner, Lyons, International Agency for Research on Cancer 1978, pp xiii and 550, Sw fr 30, distributed by the World Health Organization.

Reports on Working Groups

The Future of Mental Hospitals, Mannheim, 2–5 November 1976.

The Organization and Operation of Long-Term Care Services, Uppsala 12–15 April 1977.

Child Cancer Control, Prague 25-28 April 1977.

The Role of Health Economics in National Health Planning and Policy Making, Cologne 7–10 June 1977.

Technical Report Series

622 The Promotion and Development of Traditional Medicine. Report of a WHO Meeting

626 WHO Expert Committee on Biological Standardization 29th Report.

631 Evaluation of Certain Food Additives and Contaminants. 22nd Report of the Joint FAO/WHO Expert Committee on Food Additives.

Announcing a new Karger series

Contributions to Epidemiology and Biostatistics

Series Editor: M.A. Klingberg, Ness-Ziona/Tel-Aviv

This new series concerns the study of infectious and non-infectious diseases through the integrated use of epidemiologic and biostatistical methods. Individual volumes concentrate on a specific topic or problem with the aim of providing a compact digest of recent developments. Such works offer convenient access to information which must normally be abstracted from voluminous textbooks or scattered journal articles.

You can subscribe to this new Karger series with an introductory discount of 30%

Just published

Vol. 1: Epidemiologic Methods for Detection of Teratogens
Editors: M.A. Klingerg, Ness-Ziona/Tel-Aviv and J.A.C. Weatherall, London
X + 204 p., 22 fig., 61 tab., 1978
SFr. 83.– / DM 99.– / approx. US $ 49.75
ISBN 3–8055–2889–2

This volume covers recent progress in the understanding of how environmental factors contribute to human birth defects. Its 15 papers offer information on environmental polutants and potential occupational teratogens as well as the established teratogens, such as infectious diseases, drugs, chemicals, and radiation. Special emphasis is given to the use of epidemiologic methods for detecting teratogens in population groups.

The volume begins with a critical, up-to-date review of the established and potential teratogens in man. In subsequent papers, authors report their original work on the adverse influences of certain drugs taken during pregnancy, and on the possible occupational hazards encountered by pregnant women working in industry. Special emphasis is given to techniques for monitoring birth defects in large as well as in small populations of births. Additional topics range from a discussion of statistical methodology to a report relating birth defects to parental occupation. The volume closes with a first-time report on the prospects for international collaboration in the prevention of congenital malformations.

Future volumes

The Eradication of Infectious Diseases
A Critical Study
P. Yekutiel, Ramat-Aviv

Epidemiological Approaches to New Diseases of Unknown Etiology
I. Shigematsu, Tokyo

Epidemiology of Diabetes
J.I. Mann, Oxford

Disaster Epidemiology
J. Seaman, London

Epidemiology of Congenital Malformation
M.A. Klingberg, Ness-Ziona/Tel Aviv

Disease Clustering
N. Mantel, Bethesda, Md., and M. Klauber, Salt Lake City, Utah

S. Karger
Basel · München · Paris
London · New York · Sydney

Order Form · 30% discount

Contributions to Epidemiology and Biostatistics

_____ copy/ies of Vol. __________

_____ copy/ies of Vol __________

_____ copy/ies of Vol. __________

☐ Prospectus on the series

Name __________

Address __________

____

City/Postal code __________

____

Country __________

Please send this order form to your bookseller or to:

USA: Albert J. Phiebig, Inc., P.O. Box 352, White Plains, NY 10602 (USA)

All other countries: S. Karger AG, Arnold-Böcklin-Strasse 25, CH–4011 Basel (Switzerland)

SI 78007