Lawrence Berkeley National Laboratory
Recent Work

Title

THE EPIDEMIOLOGY OF OVARIAN CANCER IN THE UNITED STATES: AN ECOLOGICAL
ANALYSIS

Permalink

https://escholarship.org/uc/item/97n5z66v

Author

Pinto, J.A.

Publication Date

1984-10-01

Lawrence Berkeley Laboratory
UNIVERSITY OF CALIFORNIA

Computing Division

THE EPIDEMIOLOGY OF OVARIAN CANCER IN THE
UNITED STATES: AN ECOLOGICAL ANALYSIS

J.A. Pinto
(Ph.D. Thesis)

October 1984

DISCLAIMER

The Epidemiology of Ovarian Cancer in the United States: An Ecological Analysis

Jennifer Anne Pinto
(Ph.D. Thesis)

Computer Science Research Department
University of California
Lawrence Berkeley Laboratory
Berkeley, California 94720

October, 1984

ABSTRACT

There were 111,037 deaths attributed to ovarian cancer in women over age 35 (white = 102,668; non-white = 8,369) in the United States during the eleven year period 1968 - 1978. The annual rates show a dramatic increase with age, rising from about 5/100,000 in women aged 25-34 to a peak of about 125/100,000 in women aged 65 and over. Time-trend analysis revealed a slight increase in mortality from 1968 to 1978, but it is quite likely that part of this increase is due to an increased awareness and diagnosis of the disease, rather than any true increase in risk. A cohort analysis of mortality patterns from 1953-1978 indicates that women of different birth cohorts experience differing risk of ovarian cancer throughout their lifetimes, a finding that has been reported elsewhere as suggestive of an etiologic role of a woman's reproductive experience.

Ecological analysis of several hypothesized risk factors used traditional regression analysis and an alternative method which orders the population on the basis of some predictor variable of interest and partitions it into a series of equal risk categories which can then be analyzed for associations. Both methods showed birthrate in 1940 and access to medical care (as measured by degree of urbanization, socioeconomic status, and physician/population ratio) to be predictive of ovarian cancer mortality in United States counties in1970.

Ovarian cancer mortality showed no consistent geographic pattern when viewed by traditional computer mapping techniques, with the high rate counties scattered randomly throughout the country. However, when the country was divided into equal risk categories by latitude, longitude, and a latitudinal/longitudinal index representing distance from the center of the

United States, it was found that there were significantly higher rates in the central states as compared to the border states.

       This study demonstates the application of alternative techniques of ecological analysis of a large mortality database, and illustrates the value of division of the population into equal risk categories to control for a potential confounder in the analysis of geographic patterns.

I gratefully acknowledge the encouragement and inspiration I received from my adviser, Dr. Warren Winkelstein, the enthusiasm and unfailing patience of Drs. Deane Merrill and Steve Selvin, the assistance of Craig Eades, and the support of my friends.

TABLE OF CONTENTS

1. THE OVARY: BACKGROUND AND PHYSIOLOGY

1.1. Introduction

The causes of ovarian cancer are much less clearly delineated than are those of the other major gynecologic neoplasms. This lack of knowledge is distressing in light of the fact that cancer of the ovary will affect 1 to 2 percent of women in their lifetime, and will be responsible for the death of most women it does strike.[1] The annual age-adjusted incidence rate for cancer of the ovary in United States women is 14 per 100,000, almost equal to that of cervical cancer. This places it in sixth place among the most common female cancers, behind breast: 73/100,000; colon: 24/100,000; corpus: 23/100,000; cervix: 15/100,000; and lung: 14/100,000.[2] However, ovarian cancer assumes particular importance when one recognizes that among all female cancers, only lung cancer has a poorer prognosis. The median survival for all stages of ovarian cancer is 18 months.[3] Nearly 10,000 women die from this disease in the United States each year, almost as many as from all other gynecologic neoplasms combined.

The main problem in studying ovarian cancer is that it is nearly impossible to diagnose early in its development. At present, early diagnosis is a matter of chance rather than scientific method.[4] Until such methods are available, it is essential that the natural history and the epidemiology of the disease be studied.

The disciplines of epidemiology and medicine should complement

[1]Zdeb M.S. The probability of developing cancer. *Amer J Epi* 1977;106:6-10.

[2]Cutler R.T. National Cancer Institute Monograph 41: *Third National Cancer Survey Incidence Data* DHEW Pub. No. (NIH) 75-787, 1975.

[3]Axtell L.M., Asire A.J., Myers M.H. *Cancer patient survival.* Report No. 5, DHEW Pub. No. (NIH) 77-992, 1976.

[4]Barber H.R. *Ovarian Carcinoma: Etiology, Diagnosis and Treatment.* New York: Masson Pub., 1982.

one another in their attempts to reduce the burden of disease in a population. One must compare the relative importance of the forces of morbidity, mortality, and case fatality in order to identify those diseases that rely most heavily on epidemiological research for their control. Those diseases which are highly prevalent (i.e.- high morbidity rate) but for which there are effective medical procedures for diagnosis and treatment should be less of a concern for the epidemiologist than those less prevalent but highly elusive diseases that defy early medical detection and treatment, and therefore have a high case-fatality rate. Ovarian cancer is one such disease.

1.2. History

The earliest known reference to the existence of the ovaries is found in the writings of Aristotle (384 - 322 B.C.), in his description of the spaying of sows:

> The ovaries of the sows are excised with the view of quenching in them sexual appetites and of stimulating growth in size and fatness. ...They cut the lower belly about the place where the boars have their testicles, for it is there that the ovary grows, adhering to the two divisions (or horns) of the womb.[5]

The role of the ovaries in reproduction was not clearly delineated until the 19th century; the principal impediment to an earlier understanding was most likely the Aristotelean "seed-and-soil" concept of reproduction. Aristotle believed that the male was the giver of "seed", while the female provided the "soil" in which the seed could grow.

Galen (122 - 199 A.D.) added very little to Aristotle's accounts of reproduction, although he believed that the ovaries produced a kind of "sperm" that was essential to reproduction. Following the death of Galen and with the beginning of the Dark Ages, all inquiries relevant to the physiology of

---

[5]Thompson D'A.W. *The Works of Aristotle Translated Into English. Vol IV: Historia Animalium.* London/ New York: Oxford Press, 1910.

reproduction were suspended.[6]

A renewed interest in anatomy began to emerge in the 14th and 15th centuries, and scientists and anatomists began to appreciate the structural differences between the testis and the ovary. Andreas Vesalius of Brussels (1514 - 1564) and his pupil Fallopius (1534 - 1562) both made important discoveries and contributions to the anatomical knowledge of the human reproductive system.[7] The 17th century saw further advancement in the description of ovarian structure and the emergence of a heated debate about its function in human reproduction. William Harvey (1578 - 1657) produced a treatise which, while based on excellent scientific experimentation and observation, was Aristotelean in interpretation. Harvey believed that the egg was a product of, rather than a participant in conception. Other scientists were of the opinion that the female "testes" produced their own "semen", the follicular fluid, which had to mix with the semen of the male in order to give rise to an embryo. A third school of thought subscribed to the notion that the entire follicle was an egg, and that fertilization occurred within the ovary itself, only fertilized eggs being capable of escaping.

The controversy over the function of the ovary continued into the 18th century. In 1744 Boerhaave outlined the theory of how the ovum must escape from the ovary, leaving a scar or corpus luteum in the process, and pass down the Fallopian tube to be fertilized by a spermatozoon before entering the uterus.[8] The surgeon Percival Pott of London made an important discovery about the function of the ovaries in controlling menstruation, and gave the first accurate clinical description of the consequences of

[6]Short R.V. The Discovery of the Ovaries. In: Zuckerman L. and Weir B., eds. *The Ovary*. New York: Academic Press, 1977:10-96.
[7]*Ibid.*, p.15.
[8]*Ibid.*, p.18

ovariectomy.[9]

At the beginning of the 19th century two Frenchmen, Prevost and Dumas, gave a thorough account of ovulation and the formation of the corpus luteum. They realized that fertilization must occur after the egg has left the ovary, and concluded that the follicles were not the ova themselves, but were the receptacles of the ova, the follicular fluid being necessary to transport the ova into the uterus.[10] Following this line of reasoning, von Boer in 1827 first described and illustrated the mammalian egg within the follicle (Figure 1).

FIGURE 1:

OVUM

FOLLICLE

---
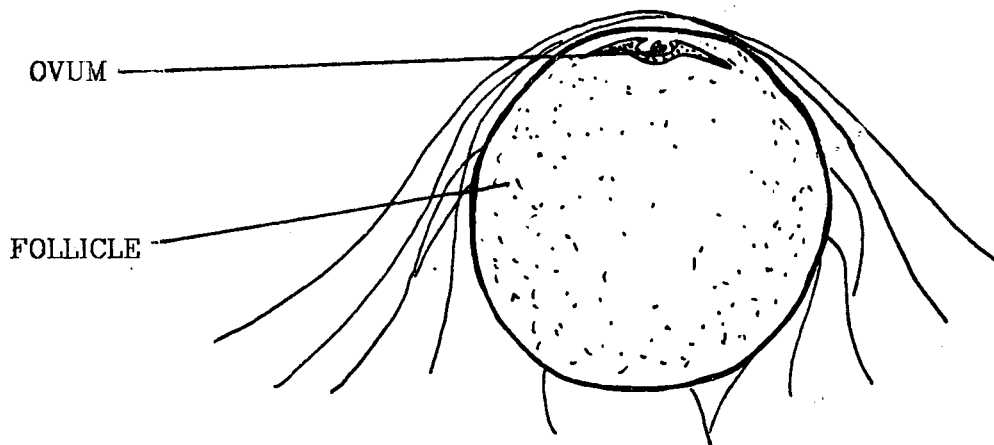
[9] Corner G.W. The relation of the ovary to the menstrual cycle: Notes on the history of a belated discovery. *Ann Fac Med* 1950; 35:758-87.

[10]Prevost J..L, Dumas J.A. De la generation dans les mammiferes at des premiers indices du developpement de l'embryon. *Ann Sci Nat* 1824; 3:113-125.

Von Boer's conclusion:

> Omne animal quod coitu maris et feminae gignatur, ex ovo evolvitur. (Every animal which is generated by coitus of male and female is evolved from an egg.)[11]

was not far removed from Harvey's conclusions two centuries before. However, the intervening years had seen the mammalian egg redefined as a component rather than a product of conception.

So, the great debate about the mechanism of conception and the function of the ovaries was finally laid to rest. The era of descriptive anatomy of the ovary was ending and the period of exploring physiology ensued. By the late 1930's, the principal steroids secreted by the human ovary, estradiol and progesterone, had been isolated, characterized and synthesized, and research into their biological effects in women continues.

1.3. Development/ Embryology

The initial development of the reproductive system in both males and females is similar. Undifferentiated gonads and two pairs of ductal systems (the Wolffian and the Mullerian) are present at about five weeks of gestation. If the fetus develops as a female, requiring two X chromosomes, the Wolffian ductal system regresses while the Mullerian ductal system develops and differentiates into Fallopian tubes, the uterus, and the upper two-thirds of the vagina. If the fetus develops as a male, due two one X and one Y sex chromosome, the Mullerian ducts regress under the influence of Mullerian inhibiting factor, and the Wolffian ducts differentiate into the male internal ductal system by the action of testosterone.

In the normal female fetus, the primordial germ cells migrate to the gonad from the yolk sac endoderm during the fifth week of gestation.

---

[11]Short op.cit., p.43.

These germ cells divide rapidly by mitosis beginning during the sixth week of gestation, and by week 20 - 24 a maximum of seven million oogonia have been produced. Oocytes are produced by mitotic division of these oogonia. By the time the infant is born all oogonia have become primary oocytes, but many have already degenerated, leaving approximately two million oocytes at birth. Of these two million, only 400 or so will ever be released through the process of ovulation.[12]

1.4. Structure and function

The mammalian ovaries are paired organs, approximately equal in size, and spherical or oval in shape. The ovary is divided into a cortical zone and a medulla and is covered by a tunica albuginea. This continuous sheet of epithelium is disrupted only following ovulation when follicles burst to release an ovum. Repair of the ruptured epithelium occurs within 2 - 4 days.[13] Each ovary contains follicles in various stages of development, ranging from the immature primordial to the fully developped Graafian. Primordial follicles consist of a single layer of flattened epithelial cells surrounding each oocyte. As they increase in size, the follicles migrate deeper into the ovarian cortex. The single layer of flattened cells enveloping the oocyte becomes cuboidal, the cells become columnar, and a distinct membrana granulosa is formed. An outer layer derived from the stroma develops as well, constituting the thecal layer. As the follicle enlarges, cavities form in the granulosa. The cavities enlarge and merge to form an antrum filled with follicular fluid.

In addition to the follicles, the ovary contains those bodies which

---

[12] Lee P.A. Ovarian function from conception to puberty: Physiology and disorders. In: Serra GB, ed. *The Ovary*. New York: Raven Press, 1983:177-89.

[13]*Ibid.*, p.179.

form after the discharge or degeneration of the ovum: corpora lutea and albicani, as well as stromal and connective tissue, interstitial tissue, and vascular, nervous and lymphatic tissue. The corpus luteum is the endocrine gland which normally develops from the cellular components of the follicle after ovulation. The corpus luteum degenerates into a corpus albicanus in the absence of fertilization and implantation of the released ovum.

There exists an intricate and delicate relationship between the pituitary and ovarian hormones in the control of ovulation. Between the ages of 2 and 8 years, circulating levels of gonadotropins and estradiol are low. The ovary is capable of steroidogenesis and ovulation when stimulated by gonadotropins, as can be seen in patients with true sexual precocity.[14] During childhood, the ovary is an active organ with follicular growth and degeneration. The pituitary and hypothalmus become less sensitive to negative feedback from the ovary during this time.[15] The first change in hormonal secretion which marks the onset of puberty is increased ovarian estrogen secretion, which results in increased growth and onset of breast development. The human menstrual cycle reflects a cyclic activity in the ovary in preparation for the possible implantation of the blastocyst in the endometrial cavity. This cycle is due to the integration of activity by the hypothalmic-pituitary axis, an interplay between the secretion of ovarian steroids and the release of gonadotropic hormones from the anterior pituitary. The synchrony necessary for fertility can take years to develop.[16]

The mature ovary responds to the cyclical endocrine stimuli with the release of a mature ovum, which passes into a Fallopian tube and travels

---

[14]*Ibid.*, p.182.

[15]*Ibid.*, p.183.

[16]*Ibid.*, p.189.

down to the uterus, the site of implantation and development of the fetus.

1.5. Classification of ovarian tumors

Ovarian tumors are as complex in histogenesis and structure as is the ovary itself. This fact, along with their anatomical position, makes the diagnosis of ovarian cancer difficult, which, in turn, affects the quality and reliability of data on ovarian cancer incidence and prevalence. Almost all of the neoplasms that arise in the ovary have their origin in one of the following: germ cells, follicular cells, or epithelial cells. The category of germ cell tumors includes of a number of rare types - mainly dysgerminomas and teratomas.[17] The category of sex-cord/stromal tumors refers to those tumors which arise in follicular cells or other cells derived from the mesenchyme of the embryonic gonad. Together, the categories of germ-cell and sex-cord/stromal tumors represent only approximately 5% of all ovarian cancers. The category of epithelial tumors refers to those tumors which arise in the ovarian epithelium. This category constitutes the large majority of ovarian cancers, representing the remaining 95%.

There are four main categories of epithelial tumors: serous, mucinous, endometriod, and clear cell. The histologic criteria used to distinguish between these ovarian tumors are highly variable, in spite of attempts to standardize the classification scheme.[18] In 1961 a committee of the International Federation of Gynecology and Obstetrics proposed a classification of epithelial ovarian tumors, and more recently a group of pathologists appointed by the World Health Organization formulated a more detailed classification, including germ cell tumors, sex-cord/stromal tumors, and oth-

---

[17]Scully R.E. *Ovarian tumors: A review*. Amer J Pathol 1970; 1:73-85.
[18]Barber H.R. *op.cit.,p.156*.

ers. A brief outline[19] of the more common groupings is presented here:

    I. Epithelial tumors
        A. Serous
        B. Mucinous
        C. Endometriod
        D. Clear cell

    II. Germ cell tumors
        A. Dysgerminomas
        B. Teratomas

    III. Sex-cord/Stromal tumors
        A. Granulosa
        B. Sertoli-Leydig

    IV. Other
        A. Gonadoblastomas
        B. Fibromas; Sarcomas
        C. Tumors not specific to the ovary: Burkitt's; Lymphoma

It is clear that interpretation of figures on the incidence of the various categories of ovarian cancers must acknowledge the important factors affecting the quality and reliability of the data. As the material accumulates and is reported by stage, grade and histologic criteria that are universally acceptable, a meaningful study can be undertaken. Meanwhile, it appears best to focus on the epidemiology of epithelial tumors which, because of their relative frequency, approximates the epidemiology of ovarian cancer as a whole.

---

[19]*Ibid.*, p.34.

2. THE EPIDEMIOLOGY OF OVARIAN CANCER: REVIEW OF THE LITERATURE

2.1. Demographic patterns

It has been observed that overall, age-adjusted mortality rates from ovarian cancer have been increasing over time. The rate of increase has not been uniform across all age groups however, and mortality rates have actually been falling in the younger age groups since the 1950's.[20] Beral has reported that in England and Wales, the downward trend in mortality at age 30-34 preceded that at age 35-39, which in turn preceded that at age 40-44.[21] Silverberg has shown that the race and age-specific mortality rates for ovarian cancer show a steady increase from 1930 until the late 1940's for all race/age groups. In recent years, the rates decreased slightly for the age group 30-44. The rates for those over age 55 in general show a slight increasing trend. Non-whites have rates lower than the whites for all ages over 30 years.[22] Age-adjusted mortality rates show a steady increase up to 1959, when the rates stabilized at 7-8/100,000.[23] Recent publications rank ovarian cancer sixth among the leading cancers in United States females, accounting for approximately 5% of all female cancer deaths.[24]

Incidence rates are much more difficult to evaluate due to the heterogeneity of histologic type and the difficulties in diagnosis. Cancer of the ovary is one of the most inaccessible, and confirmation of diagnosis of

[20]World Health Organization. *World Health Statistics Annual, 1955-1975.* W.H.O., Geneva, 1975.

[21]Beral V. The Epidemiology of Ovarian Cancer. In: Newman C.E., Ford C.H.J. and Jordan J.A., eds. *Ovarian Cancer.* Pergamon Press, New York, 1979.

[22]Silverberg E. *Statistical and Epidemiological Information on Gynecologic Cancer.* American Cancer Society Pro. Ed. Pubs., New York, 1980.

[23]Wynder E.L., Dodo H., Barber H.L. Epidemiology of cancer of the ovary. *Cancer* 23:352, 1969.

[24]Barber H. *op.cit.*

the disease requires post-mortem evaluation.[25] Part of the reported increase in ovarian cancer in the United States may be attributable to improved diagnostic facilities and techniques and increased awareness of the disease. Furthermore, ovarian cancer may be more often diagnosed in individuals to whom better medical care is available, making incidence rates broken down by race, socioeconomic status and education subject to ascertainment bias.

As already indicated, there are three main categories of ovarian tumors, classified by the cell type of origin: germ-cell tumors, sex cord-stromal cell tumors, and epithelial cell tumors. The category of germ cell tumors is exceedingly rare. Based on current cross- sectional data, their incidence peaks in young adulthood, falls, rises once again, and levels off after middle age. At no age does the annual incidence exceed 1.0 per 100,000.[26] Because of the rarity of these tumors and the fact that population based registries have only recently begun to report data by histologic type, time trends in incidence have never been evaluated. Mortality from germ cell tumors has been analyzed for the period 1950-70, and appears to have remained relatively constant.[27]

Sex cord-stromal tumors are also very rare. In the United States the incidence rises with age, to reach a maximum of .5 to 1.0 per 100,000 per year. As with germ cell tumors, there are no time-trend incidence data available. The incidence of both germ cell and sex cord- stromal tumors appears to be similar among the major racial and ethnic groups in the United States according to the one study that examined this question.[28] However, because

[25]*Ibid.*, p.27.

[26]Weiss N.S. The Ovary. In: Schottenfeld D., Fraumeni J.F., eds. *Cancer Epidemiology and Prevention.* W.B.Saunders Co.: New York, 1982.

[27]Li F.P., Fraumeni J.F., Dalager N. Ovarian cancers in the young: Epidemiologic observations. *Cancer* 32:969, 1973.

[28]Weiss, *op.cit.*, p.871.

of the potential ascertainment bias mentioned previously, and the small number of cases identified in this study, no conclusive statements can be made.

Epithelial tumors constitute the majority of ovarian cancers, representing 9% of all. Again, the incidence rises with age, reaching a maximum of 35-40 per 100,000 per year. Although the epidemiology of epithelial ovarian cancer is more clearly delineated than that of either germ cell or sex cord-stromal tumors, problems in ascertainment and diagnosis still exist. These tumors can be difficult to diagnose, particularly when the malignancy is widespread, and in many women with advanced abdominal malignancy, the diagnosis of ovarian cancer is only an educated guess.[29] At the other end of the spectrum, tumors that do not appear to have yet invaded the stroma are labelled borderline, and are not included in most cancer statistics.[30] Furthermore, the determination of the specific histologic sub-type among epithelial tumors is subject to variation. The internationally accepted International Federation of Gynecology and Obstetrics classification scheme (See Chapter 1, Section 5) which was devised less than twenty years ago is still not consistently employed by pathologists.

The incidence of epithelial tumors is very low in pre-pubertal females, rising rapidly after puberty until the fifth and sixth decades of life, when it levels off.[31] Weiss et.al.[32] used Third National Cancer Survey (TNCS) data 1968-71 to evaluate the relative incidence of histologic types: epithelial versus nonepithelial (germ cell and sex cord-stromal) tumors. Their results

[29]*Ibid.*, p.872.

[30]Wynder E.L., Dodo H., Barber H.R. *op.cit.*, p.355.

[31]Weiss, *op.cit.*, p.873.

[32]Weiss N.S., Homonchuck T., Young J.L. Incidence of histologic types of ovarian cancer: The U.S. TNCS, 1969-71. *Gynecol Oncol* 1977a;5:161-85.

are presented in Table 1 below.

TABLE 1:			
Incidence of Ovarian Cancer by Histologic Type: TNCS 1968-1971			
Histologic type	Age	Incidence (per 100,000)	
		White	Black
Epithelial			
	<40	1.4	1.4
	40-59	26.0	17.7
	>60	38.4	24.3
Non-epithelial			
	<40	0.3	0.5
	40-59	0.9	1.1
	>60	1.2	1.2

One study examined incidence data from the TNCS program, 1969-1971, and the Surveillance, Epidemiology and End Results (SEER) program, 1975-77, (using four common areas: Atlanta, Detroit, Iowa, and San Francisco/Oakland) and reported a significant increase in the incidence of endometrioid and clear-cell cancers of the ovary, while the overall incidence remained relatively stable. The authors concluded that while a shift in criteria for histologic classification might explain part of the increase (rates for unspecified epithelial tumors declined during the same time period), it is unlikely that it accounts for all of it. They suggest that the concomitant increase in cancer of the uterine cervix suggests a common etiologic factor, perhaps use of post-menopausal estrogens.[33]

Morbidity and mortality rates for ovarian cancer vary widely by country. Segi reports the following annual age-adjusted mortality rates: Denmark: 11.2, Sweden: 9.2, Norway: 8.3, England and Wales: 8.0, Scotland: 7.9, New Zealand 7.7, Canada: 7.2, United States(white): 7.0, Israel: 6.6, South Africa: 6.4, Australia: 6.1, Finland: 5.8, United States(non-white): 5.6, Bel-

---

[33]Cramer D.W., Devesa S.S., Wetch W.R. Trends in the incidence of endometrioid and clear-cell cancers of the ovary in the Unites States. *Am J Epidemiol* 1981;114:201-223.

gium: 5.5, Ireland: 5.0, Italy: 3.0, Japan: 1.7.[34] Mortality rates increased in most countries between the years 1950 and 1962, but these increases may reflect improved diagnostic techniques and awareness, rather than an actual increase in incidence. The incidence rates in various countries have been analyzed by the International Union Against Cancer, and there exists a seven-fold difference between the country with the highest annual incidence rate, Sweden: 21/100,000 and the country with the lowest incidence rate, Japan: 3/100,000.[35] The difference in incidence rates may in part be explained by the difference in age distributions among the various populations. However, further analyses reveal that this is not the only explanantion; the patterns persist when age-adjusted rates are compared.

Mortality rates for ovarian cancer in Japanese migrants to Hawaii are significantly higher than those in Japan. This pattern persists for their first generation offspring as well.[36] [37] In fact, rates in these migrants are intermediate to the overall rates of the countries of origin and residence. Such findings would tend to implicate environmental and dietary influences, however, it must be remembered that the migrants are not usually a representative sample of the habitants of the country of origin.

Several researchers have hypothesized that covariation in absolute rates of certain cancers by geographic area might be due to common etiologic factors. In the United States, age-adjusted mortality by cancer site was mapped at the county level for the years 1950-69. The authors present sug-

[34]Segi M., Kurihara M.M. *Cancer Mortality for Selected Sites in 24 Countries. No.4.* Tohoku University School of Medicine, Sendai, Japan, 1966.

[35]Doll R., Muir C., Waterhouse J., eds. *Cancer Incidence in 5 continents, Vol. 2.* Springer-Verlag, Berlin, 1972.

[36]Buell P., Dunn, J.E. Cancer mortality among the Japanese Isei and Nisei of California. *Cancer* 1965;18: 656-678.

[37]Haenszel W., Kurihara M. Studies of Japanese migrants. I: Mortality from cancer and other causes. *JNCI* 1968;40:43-67.

gestive evidence of a correlation between ovarian and breast cancers by the striking similarity in the pictorial representation of the mortality rates of these two cancer sites.[38] Incidence data from the Third National Cancer Survey was used iby separate investigators to evaluate the geographic variation in the occurrence of several cancer sites. Variation was measured by product moment correlation coefficients to summarize the association between pairs of cancers. A positive correlation ($r > .50$) was reported for the female sexual sites breast, corpus and ovary, lending further support to the hypothesis of common etiologic factors.[39]

While the variation in incidence and mortality is striking, the lack of standardized terminology and diagnostic criteria, and the international variation in availability of diagnostic and treatment facilities render comparisons difficult to interpret and possibly invalid. A more reasonable approach to understanding the etiology of ovarian carcinoma appears to be to study its distribution and variation within one country.

2.2. Etiologic factors

2.2.1. Host factors

2.2.1.1. Endocrinologic and gynecologic

Hormonal malfunction as measured by menstrual history has been demonstrated to be an important predictor variable in two matched case-

---

[38]Hoover R., Mason T.J., McKay F.W., et.al. Geographic patterns of cancer mortality in the United States. *In:* Fraumeni J.F. (ed): *Persons at High Risk of Cancer.* New York, Academic Press, 1975, p.343-60.

[39]Winkelstein W., Sacks S.T., Ernster V.L., Selvin S. Correlations of incidence rates for selected cancers in the nine areas of the Third National Cancer Survey *Am J Epidemiol* 1977;105:407-419.

[40]West R.O. Epidemiologic study of malignancies of the ovaries. *Cancer* 1966;19:1001-1007

control studies of ovarian cancer.[40] [41] These results are difficult to evaluate because of the various histologic types included in the studies and the reliance on patient recall for information concerning reproductive and menstrual history and endocrine profile. It is possible that patients with ovarian cancer may exhibit a bias in recall of menstrual history. Specifically, these women may be more likely than controls to report a history of dysmenorrhea or abnormal menstrual bleeding. Evidence supporting an etiologic role of hormonal disorders in ovarian cancer include: 1) the rarity of epithelial tumors before menarche and in the oldest age groups, 2) the dramatic increase in rates observed at the extremes of reproductive life when ovarian function is beginning or ending, and 3) the sharing of host characteristics with breast cancer.[42] Observations of women with multiple primary neoplasms add to the evidence that cancer of the ovary shares a similar etiology to that of cancer of the breast.[43] [44] Women with cancer of the breast have twice the expected risk of subsequently developing a separate cancer of the ovary, and women with ovarian cancer are three to four times as likely as those without it to develop breast cancer.[45]

2.2.1.2. Reproductive experience

Until recently, there was inconclusive evidence on whether decreased parity itself actually increased the risk of ovarian cancer - two

[41]Wynder E.L., Dodo H., Barber H.R. op.cit., p.367.

[42]Lingeman C. Etiology of cancer of the human ovary: A review. JNCI 1974;51:1603-1618.

[43]Schottenfeld D., Berg J. Incidence of multiple primary cancers. IV: Cancer of the female breast and genital organs. JNCI 1971;46:161-67.

[44]Schoenberg B.S., Greenberg R.A., Eisenberg H. Occurrence of certain multiple primary cancers in females. J Natl Cancer Inst 1969;43: 15-24.

[45]Ibid., p.25.

[46]Stewart H.L., Dunham L.J., Casper J., et.al. Epidemiology of cancers of the uterine cervix and corpus, breast and ovary in Israel and New York. J Natl Cancer Inst 1966;37:1-18.

studies reported an effect[46] [47] and two studies reported no effect.[48] [49] In the late 1970's, results from seven additional case-control studies on this issue were published, all supporting the notion that increased parity is protective.[50] [51] [52] [53] [54] [55] [56]

There are two possible explanations for this finding: pregnancy might actually afford some protection from ovarian cancer, or, low parity and ovarian cancer might both be related to a confounding variable which underlies both conditions, such as an endocrinologic or hormonal disorder. Beral[57] argues that pregnancy itself is protective by demonstrating that mortality rates from ovarian cancer correlate inversely with average completed family size over time. Beral also argues that the variation in average completed family size might explain much of the observed international variation in ovarian cancer rates.

Fathalla[58] has proposed a physiological model to explain the observed protective effect of pregnancy. He observes that women experience "purposeless ovulation" almost continuously from puberty to menopause. The

[47]Joly D.J., Lilienfeld A.M., Diamond E.L., et.al. An epidemiologic study of the relationship of reproductive experience to cancer of the ovary. *Am J Epidemiol* 1974;99:190-209.

[48]West, *op.cit.*

[49]Wynder, *op.cit.*

[50]Annegers J.F., Strom H., Decker D.G., et.al. Ovarian cancer: Incidence and case-control study. *Cancer* 1979;43:723-729.

[51]Casagrande J.T., Louie E.W., Pike M.C., et.al. Incessant ovualtion and ovarian cancer. *Lancet* 1979;II:170-172.

[52]Demopoulos R.I., Seltzer V., Dubin N., et.al. The association of parity and marital status with the development of ovarian carcinoma: Clinical implications. *Obstet Gynecol* 1979;54:150-169.

[53]Lau M.H., Petschelt E., Poehls H., et.al. Epidemiology of ovarian carcinoma. *Arch Geschwulstforsch* i1977;47:57-74.

[54]McGowan L., Parent L., Lednar W., et.al. The woman at risk for developing ovarian cancer. *Gynecol Oncol* 1979;7:325-33.

[55]Newhouse M.L., Pearson R.M., Fullerton J.M., et.al. A case-control study of carcinoma of the ovary. *Br J Prev Soc Med* 1977;31:148-154.

[56]Wolnik L., Bauer H. Epidemiology of ovarian cancer. *Onkologie* 1979;2:96-110.

[57]Beral, *op.cit.*

[58]Fathalla M.F. Incessant ovulation - A factor in ovarian neoplasia? *Lancet* 1971;I:163-73.

ovarian surface epithelium is subject to almost constant stimulation, and epithelial inclusions at the site of ovulation might be the sight of neoplastic growth.[59] Under this model, pregnancy would confer protection, as would any other factor which inhibited ovualtion. Consistent with this is the finding that oral contraceptive use decreases the risk of ovarian cancer.[60] [61]

The results of a convincing and well done study of this hypothesis were published in 1979.[62] In this investigation, three factors: number of live births, number of incompleted pregnancies, and oral contraceptive use were combined into a single index to produce a measure of anovulatory periods, or "protected time". This index was shown to vary inversely with ovarian cancer risk (epithelial tumors, specifically). Since the incidence of ovarian epithelial tumors rises exponentially with age, the logarithm of ovarian age (OA) was considered to be an appropriate indicator of risk. The logarithm of OA, with ovarian age defined as the period from menarche to menopause, or the time of ovarian cancer diagnosis minus "protected time", was strongly associated with ovarian cancer risk throughout its distribution.

2.2.1.3. Familial and genetic

It appears that women with ovarian cancer are more likely than controls to have relatives with the disease. Several reports describe families in which there exists familial aggregation of the disease, but because of the small numbers involved, it is not possible to rule out chance as an explanation. Case-control studies should be able to determine whether there exists familial aggregation, but the change in diagnostic practices across the

---

[59]Zajicek J. Prevention of ovarian cystomas by inhibition of ovulation: A new concept. *J Reprod Med* 1978;20:114-122.

[60]Newhouse, *op.cit.*

[61]Casagrande *op.cit*

[62]Casagrande, *op.cit.*

generations limits the sensitivity of such studies. Of the three case-control studies that have examined this issue, two found no difference in the family history of ovarian cancer patients and their controls[63] [64] and the third found an increased frequency of ovarian cancer in the maternal relatives of the cases.[65]

Unusual susceptibility to ovarian cancer occurs in females with two rare syndromes: Peutz-Jeghers and basal-cell nevus syndromes.[66] [67] Both syndromes appear to be inherited due to autosomal dominants.[68] Lymphocytes of patients with the most common types of ovarian cancers have not shown consistent chromosomal abnormalities. Chromosomes of cells of benign tumors are diploid, whereas those from malignant tumors are multiploid.[69] The highest ploidy is associated with the most aggressively malignant neoplasia.[70] Attempts to identify genetic markers have been unsuccesful to date, although one study reported that ovarian cancer is more common in women of blood group A.[71] These results must be confirmed elsewhere.

2.2.1.4. Other factors

As might be expected, single women have higher rates of ovarian

[63]Lau, op.cit.

[64]Wynder, op.cit.

[65]Casagrande, op.cit.

[66]Christian C.D. Ovarian tumors: An extension of the Peutz-Jeghers syndrome. Am J Obstet Gynecol 1971;111:529-532.

[67]Berlin N.I. Basal cell nevus syndrome. Ann Int Med 1966;64:403-415.

[68]Lingeman, op.cit.

[69]Atkin N.B., Baker N.C. Chromosomal abnormalities as primary events in human malignant disease: Evidence from marker chromosomes. J Natl Cancer Inst 1966;36:359-364.

[70]Atkin N.B. Modal DNA value and chromosome number in ovarian neoplasia. Cancer 1971;27:1064-1075.

[71]Osborne R.H., DeGeorge F.V. The ABO blood group and neoplastic disease of the ovary. Am J Human Genetics 1963;15:380-393.

cancer than married women, probably due to their low parity.[72] [73] The higher rates of ovarian cancer in single women have been found for tumors of epithelial origin, but not for germ cell or sex cord-mesenchyme tumors.[74] Rates of ovarian cancer have been shown repeatedly to be directly associated with socioeconomic status, or social class (i.e.- women of higher SES status have higher rates).[75] [76] [77] However, this observation may be partially due to an underlying ascertainment bias: women of higher socioeconomic classes are more likely to have their ovarian cancer detected and diagnosed than those of lower social classes. The observation that white women in the United States have ovarian cancer rates nearly twice those of non-white women[78] [79] [80] may also reflect this difference in ascertainment.

One study in New York City revealed that Catholic women have low rates of ovarian cancer, Protestant women have intermediate rates, and Jewish women have high rates.[81] The fact that this pattern represents the inverse of their parity, Jewish women in the United States having, on the average, small families and Catholic women having larger families,[82] lends support to the parity hypothesis.

[72]Ernster V.L., Sacks S.T., Selvin S., et.al. Cancer incidence by marital status: U.S. Third National Cancer Survey. *J Natl Cancer Inst* 1979;63:567-573.

[73]Demopoulos, *op.cit.*

[74]Weiss, N.S., Young J.L., Roth G.J. Marital status and ovarian cancer in the Third National Cancer Survey. *J Natl Cancer Inst* 1977;58:913-932.

[75]Cohart, *op.cit.*

[76]Graham S., Levin M, Lilienfeld A.M. The Socioeconomic distribution of cancer of variuos sites in Buffalo, New York, 1948-1952. *Cancer* 1960;13:180-187.

[77]Wynder, *op.cit.*

[78]Barber, *op.cit.*

[79]Silverberg, *op.cit.*

[80]Weiss.

[81]McMahon B. The ethnic distribution of cancer mortality in New York city. *Acta Int Cancer Congress* 1955;16:1716-1725.

[82]Beral, *op.cit.*

2.2.2. Environmental factors

2.2.2.1. Irradiation

Two studies found an excess of ovarian cancer among women who had been irradiated for treatment of benign pelvic conditions.[83] [84] The excess number of cases observed was very small: 7 observed versus 3.8 expected in the first study, and 4 observed versus 3.1 expected in the second, and the possibility of some antecedent ovarian condition was not ruled out in either study, rendering the results less than conclusive. A case-control study desigend to examine this question observed a similarly modest elevation in risk.[85] Data from the Japanese survivors of the atomic bomb suggest that irradiation is an etiologic factor for ovarian cancer.[86] It seems that while high doses may induce ovarian cancer in women, irradiation is not a significant risk factor.

2.2.2.2. Chemical carcinogens

Two chemical agents have been studied as etiologic agents in the development of ovarian cancer: asbestos and talc. A slight excess in mortality from ovarian cancer was found in one study of industrial workers exposed to asbestos in large quantities.[87] One study found that microscopic talc particles are more prevalent in ovarian and other gynecologic tumors than in nor-

[83]Doll R., Smith P.G. The long-term effects of X-irradiation in patients treated for metropathia hemorrhagica. *Br J Radiol* 1968;41:362-367.

[84]Stander R.W. Irradiation castration: A follow-up of results in benign pelvic disease. *Obstet Gynecol* 1957;10:223-233.

[85]Annegers J.F., Strom H., Decker D.G., et.al. *op.cit.*

[86]Beebe G.W., Kato H., Land C.E. *Mortality experience of atomic bomb survivors, 1950-1974, Life Span Study. Radiation Effects Research Foundation, TR1-77, 1977.*

[87]Newhouse M.L., Berry G., Wagner J.C. A study of the mortlity of female asbestos workers. *Br J Industr Med* 1979;29:134-139.

mal tissue.[88] These findings have not been confirmed, in epidemiological studies nor in animal studies.

2.2.2.3. Infectious agents

There is no epidemiologic or experimental evidence to support the hypothesis that ovarian cancer is viral in origin. Attempts to culture viruses from ovarian tumor cells have failed.[89] Mumps virus infection, which can affect the ovaries, has been shown repeatedly to be less common among ovarian cancer patients than controls.[90] [91] [92] [93]

2.2.2.4. Exogenous hormones

Since oral contraceptives inhibit follicle growth and ovulation, these agents might be expected to have a protective effect similar to that of pregnancy on the risk of ovarian cancer. Two case-contol studies have examined this issue, and both found a significant reduction in risk of ovarian cancer among oral contraceptive users.[94] [95] However, the number of users in both studies was small, and confidence intervals around the reported relative protection were large. These findings should be interpreted with caution until results from other epidemiologic studies are available.

---

[88]Henderson W.J., Joslin C., Turnvull A.C., et.al. Talc and carcinoma of the ovary and cervix. *J Obstet Gynecol Br Comm* 1971;78:266-278.

[89]Lingeman, *op.cit.*

[90]Menczer J., Modan M., Ranon L., et.al. Possible role of mumps virus in the etiology of ovarian cancer. *Cancer* 1979;43:1375-1383.

[91]Newhouse, *op.cit.*

[92]Wynder, *op.cit.*

[93]West, *op.cit.*

[94]Newhouse, *op.cit.*

[95]Casagrande, *op.cit.*

The effect of non-contraceptive estrogens has not been well studied to date. One study found a slight excess of ovarian cancer among women who had taken cojugated estrogens for six months or more.[96] A second study found no such effect.[97] Clearly, further research on the effects of exogenous hormones on the risk of ovarian cancer is needed.

---

[96]Hoover R., Gray L.A., Fraumeni J.F. Stilbesterol and the risk of ovarian cancer. *Lancet* 1977;2:533-543.

[97]Annegers, *op.cit.*

3. MORTALITY FROM OVARIAN CANCER: AN ECOLOGICAL INVESTIGATION

3.1. Use of mortality data

There is much discussion in the cancer epidemiology literature on the relative advantages and disadvantages of mortality data and incidence data in cancer research. Ascertainment bias may render incidence data less representative than mortality data in that certain socio-demographic variables may influence the timing and likelihood of diagnosis. This may result in a biased or non-representative study population. On the other hand, mortality data may mask an underlying pattern in a particular subgroup, e.g. - earlier onset of disease in a certain racial group. However, for some cancers, the survival time following diagnosis is so short that there is very little distinction between incidence and mortality data. Ovarian cancer, with the problems in early detection and the subsequent high case-fatality rate, is one example.

This point is well illustrated by Ries et.al. in their paper on cancer patient survival.[98] Using data from the SEER program, 1973-1979, they calculated observed and relative survival* rates from 368,263 cases of first primary cancers.

[* The relative survival rate (RSR) is the ratio of the observed survival for a particular age, sex, race patient group to the expected survival for a similar age, sex, race group in the general population. In other words, it estimates the chance of surviving the effect of a particular type of cancer.]

Of 14 sites: stomach, colon, rectum, pancreas, lung and bronchus, melanomas of the skin, breast, cervix uteri, corpus uteri, ovary, bladder, kidney,

[98]Ries L.G., Pollack E.S., Young J.L. Cancer patient survival: the Surveillance, Epidemiology and End Results Program, 1973-79. Biometry Branch, National Cancer Institute, National Institutes of Health, Bethesda, MD.

brain, and non-Hodgkins lymphoma, ovarian cancer ranked tenth in terms of both observed and relative survival rates, with only 31% (observed) or 34% (relative) surviving 5 years from the date of diagnosis. Only pancreatic cancer, lung cancer, stomach cancer and brain cancer showed a poorer prognosis. Given the low relative survival rate (i.e. - high case fatality) and the numerous difficulties in diagnosis and classification of ovarian tumors, the analyses were conducted using the mortality data described below.

3.2. Description of data

3.2.1. Socio-Economic-Environmental-Demographic Information System

The Department of Computer Science Research at the Lawrence Berkeley Laboratory is responsible for the development of an interactive data retrieval system: the Socio-Economic-Environmental-Demographic-Information System (SEEDIS). SEEDIS provides interactive access to an extensive collection of databases including Census data, mortality data, and cancer incidence data. It includes such features as: selection of geographic level of analysis (e.g.: census tract, county, state, or nation) and scope of analysis (e.g.: specific tracts within one county, all states in the United States, etc.); and facilities for graphic analysis and display of data (e.g.: mapping, charts, tables, graphs). The relative ease with which one can gain access to these very large databases containing important epidemiological information makes this an important tool for epidemiologists interested in ecological studies. Following is a list of the databases from SEEDIS used in this study:

1. 1950-1980 Population by Age, Race, and Sex
2. United States Cancer Mortality, 1950-1969 (Age-adjusted)
3. 1968-1978 Cause/ Age/ Sex/ Race Mortality
4. 1947-1977 City County Data Book
5. Areas, Centroids, and Boundaries

3.2.2. Population data

3.2.2.1. Source

The population data were acquired from the Environmental Protection Agency and are based on the April 1 census counts for 1950, 1960, 1970, and 1980. It has long been recognized that artifacts in mortality trends may result from errors in the population census, taken every decade in the United States. One author has estimated the net census undercount of the population by age, sex, and race in the 1970 Census, and demonstrates that the degree of error differs by these characteristics.[99] This fact, combined with variation in the completeness of death certification by these same characteristics, makes the evaluation of mortality trends a complex task.

This issue is further complicated by the need to estimate the growth of the population across intercensal time periods. Typically, a linear interpolation method is employed, using two census values to derive an intercensal estimate. The resulting values are labelled, erroneously, the estimated midyear population. The midyear population, in fact, would be represented by a July 1 population count.

3.2.2.2. Interpolation program

In order to investigate uncertainties due to the different interpolation procedures and in hopes of generating a more accurate estimate for the intercensal years, a Fortran program was written (See Appendix; Item A) which interpolates the census values according to one of three methods:

1. Staight-line interpolation
2. Polynomial interpolation
3. Logarithmic polynomial interpolation (i.e.- polynomial interpolation of the logarithm of the population)

---

[99]Siegel J.S. Estimates of the coverage of the population by sex, race, and age in the 1970 Census. *Demography* 1974;11:1-23.

Census values from April 1, 1950, 1960, 1970, and 1980 are used to calculate an age-race-sex specific count for any geographic area of interest, for any year from 1950 to 1980, and any one of four dates: January 1, April 1, July 1 or October 1. Extrapolated values, although unreliable, are also available.

There are advantages and disadvantages to each of these methods. The advantage of the linear method is that it will lead to values consistent with the majority of published values, as this is the method most often employed. The disadvantage of this method is that because it uses only two points in time to derive an interpolated value (e.g.- 1950 and 1960 population counts would be used to derive a 1955 estimate), there is discontinuity in the population growth curve, which may be misinterpreted. The polynomial method is continuous, employing all four census values in its interpolation, and resulting in a smooth curve of population growth across time. In addition, both of these methods (linear and polynomial) have the advantage of being additive; that is, interpolated county values sum to the interpolated state value, or interpolated state values sum to the interpolated nation value. Furthermore, both of these methods generate estimates for April 1, 1950, 1960, 1970 and 1980 that are identical to the observed values. The logarithmic polynomial method has the advantage that it can never result in a negative value for a population interpolation. However, it does not have the additive property of the other two methods. The polynomial method with its advantage of additivity and the fact that it makes use of all four censal values in its estimates, appears to be the most reliable method of interpolation. After investigating the differences between the three methods, it was decided to use the polynomial method for all population estimates. estimates.

3.2.3. Mortality data

3.2.3.1. Source

The age-adjusted mortality data were acquired from the National Cancer Institute and are the same data used to produce the Atlas of Cancer Mortality, 1950-1969.[100] The data are available for any county or county group (as defined by the National Cancer Institute), by sex and race (white, non-white). The deaths are tabulated for 35 groups of diseases using the International Classification of Disease (7th Revision) codes. The age-specific and cause-specific mortality data were developed for the Mortality Surveillance Project (MSP) of Johns Hopkins University from some 20 million death certificate records from the National Center for Health Statistics.[101] This data tape contains cause-specific death counts (International Classification of Disease 8th Revision), county (as defined by the Johns Hopkins MSP), age (by 10 year age groupings), sex, and race (white, non-white) for the years 1968-1978. This tape was received by LBL in April, 1983.

3.2.3.2. Calculation of rates

For the purposes of this analysis, the number of deaths for the years 1968-1978 for white females and non-white females were extracted for the following age groupings: Total, 35-44, 45-54, 55-64, 65-74, 75-84, 85+. These death counts were used with the appropriate denominators (derived by the population interpolation program using the polynomial method) to calculate age-specific rates. The same procedure was used for both state and county rates. Crude rates for the United States were derived by aggregation

---

[100]Mason T.J., McKay F.W., Hoover R. Atlas of Cancer Mortality for United States Counties: 1950-1969. DHEW Pub. No. 75-78; NCI, NIH.

[101]Gittelsohn A., Diener M., Mead L., et.al. Notes on a national mortality system - A preliminary report. Dept. of Biostatistics; Johns Hopkins School of Hygiene and Public Health.

of the state level data.

3.2.3.3. Use of truncated crude rates versus age-adjusted rates

Because differences in the age distribution of populations within specific geographic areas or during certain time periods have a marked effect on the mortality rates in those populations, it is often desirable to remove the effect of the differences in age distribution by standardization of mortality rates. The crude death rate is actually a weighted average of the age-specific death rates in which the numbers or proportions in each age group are the weights. If population X in 1978 has a higher proportion (i.e.-increased weighting) of older persons for whom age-specific rates are higher than does population X in 1968, the crude death rate will be higher, even if the risk of dying in each age group is the same. The age adjusted rate takes into account the differences in the age distribution of the population and produces a summary index of mortality experience. This index can be used to compare the mortality experience by geographic area or year while controlling for the effect of any differences in age distributions.

Age-adjusted mortality rates by state were calculated for the years 1968-1978. The direct method of age adjustment was employed. In the direct method, the adjusted rates are derived by applying the age specific rates of the study areas to the age structure of some standard population, in this case, the United States female population, white and non- white, in 1950.

The difference between the age-specific (crude) and the age-adjusted mortality rates for each state result from the differential proportion of older women for whom ovarian cancer mortality rates are the highest. The differences between the total crude rates and the age-adjusted rates are not dramatic, but these differences can be further minimized by calculating

a "truncated" crude rate based only on the population of interest, i.e.-women aged >35. The average annual truncated crude rates are presented in Table 2A. The reliability of the truncated rate depends on the number of cases and the magnitude of the rate. The ratio of the standard error of each rate to the rate itself was calculated, and all those for which the standard error is >5% of the rate are marked by an asterisk to caution the reader that these rates are subject to substantial random variation. Table 2B shows the average annual difference (1968-1978) by state between the age-adjusted rate and the total crude rate (Column A), and the age-adjusted rate and the truncated crude rate (Column B) for the United States and each state. As can be seen, the process of truncating minimizes the disparity between the crude and adjusted rates, rendering the difference negligible.

TABLE 2A:					
Ovarian cancer: Average annual crude mortality rate per 100,000 (truncated) United States; White and non-white females; 1968-1978.					
State	White	Non-white	State	White	Non-white
AL	7.9	6.4	MT	7.8	2.6*
AK	3.5	1.5	NE	11.7	6.1
AZ	8.1	2.9	NV	7.7	2.2*
AR	8.8	6.3	NH	11.0	3.2
CA	10.1	4.0	NJ	12.0	5.9
CO	8.1	1.9	NM	7.0	1.8
CT	10.7	3.5	NY	12.0	4.3
DE	9.4	3.7	NC	7.6	4.1
DC	19.3	6.1	ND	10.6	3.5
FL	12.2	5.6	OH	10.1	5.0
GA	7.5	5.1	OK	8.9	6.0*
HI	4.4	4.9	OR	10.2	5.1
ID	7.7	4.8	PA	10.8	6.2
IL	11.2	5.1	RI	10.9*	5.7
IN	9.1	4.6	SC	7.2	4.2
IA	11.3	3.3*	SD	11.3	6.3
KS	10.0	4.9	TN	8.6	6.4
KY	8.8	9.0	TX	7.8	5.8*
LA	6.0	4.0	UT	8.1	0.3
ME	10.9	4.9*	VT	9.7	1.3*
MD	8.8	5.2	VA	8.1	5.6
MA	11.1	4.0	WA	9.8	4.5
MI	9.1	5.4	WV	8.9	9.5*
MN	10.1	1.8	WI	10.9	2.9
MS	7.0	5.5	WY	7.3	1.6
MO	10.7	6.9			

* Indicates standard deviation >5% of the rate.

TABLE 2B:								
Ovarian cancer mortality rates per 100,000; Average annual age-adjusted mortality rates minus crude rates (Column A) and age-adjusted mortality rates minus truncated rates (Column B); United States; White and Non-white females; 1968-1978.								

	COLUMN A		COLUMN B			COLUMN A		COLUMN B	
STATE	WF	NF	WF	NF	STATE	WF	NF	WF	NF
AL	-2.29	-2.01	-1.79	-1.46	MT	-1.00	-.23	-.46	.49
AK	-2.71	-1.00	-1.44	-1.12	NE	-1.84	1.13	-.94	2.12
AZ	-2.21	-.19	-1.66	.24	NV	-.66	.28	-.07	.46
AR	-2.59	-1.44	-1.01	-.94	NH	-.36	1.67	.41	2.03
CA	-.72	-.27	-.02	.25	NJ	-1.68	.0	-.91	.62
CO	-.13	.56	.46	1.12	NM	-2.05	-.04	-1.64	.01
CT	-1.05	.47	-.28	.83	NY	-1.69	-.55	-.94	.12
DE	-.43	.83	.09	1.14	NC	-2.04	-1.44	-1.54	-1.03
DC	-7.15	-1.06	-1.98	-.49	ND	-2.46	.79	-1.73	1.12
FL	-3.56	-.41	-1.78	.24	OH	-.77	-.64	-.06	-.14
GA	-1.47	-.88	-.95	-.36	OK	-2.04	-.82	-1.41	-.02
HI	-3.69	-4.27	-1.22	-1.73	OR	-1.36	-.25	-.62	.50
ID	-1.63	.30	-1.11	1.58	PA	-2.15	-1.01	-1.43	-.49
IN	-1.09	.32	-.36	.84	RI	-1.82	.59	-1.16	1.11
IA	-.73	-.46	-.05	.13	SC	-1.92	-1.36	-1.34	-.84
KS	-1.37	1.33	-.63	1.69	SD	-2.70	.67	-1.96	1.71
KY	-1.51	.42	-.87	1.03	TN	-2.16	-.94	-1.55	-.46
LA	-1.86	.14	-1.25	.72	TX	-1.50	-.30	-.93	.36
ME	-1.07	-.44	-.62	-.13	UT	-.80	.07	-.33	-.03
MD	-1.61	.77	-.84	1.37	VT	-.44	.0	.13	.0
MA	-1.03	-.27	-.34	.24	VA	-1.46	-1.06	-.88	-.61
MI	-.89	.40	-.17	.96	WV	-2.83	-3.96	-1.20	-1.26
MN	-1.06	-.86	-.44	-.35	WI	-1.23	.22	-.50	.50
MS	-1.11	.46	-.42	.58	WY	-1.65	.18	-.99	.38
MO	-1.50	-1.44	-1.08	-.96	US	1.7	.33	1.0	.20

At the county level, the study populations are too small to allow for use of the direct method of age adjustment. In this situation, the indirect method is usually employed, and rates from a standard population are applied to the age distribution of the study populations. This technique yields the expected number of deaths one would expect to see if the study population experienced the mortality rate of the standard population. The total number of expected deaths is calculated and compared to the total number of observed deaths and expressed as a Standard Mortality Ratio (SMR).

[SMR = observed deaths/expected deaths]

The Standard Score (SS) used in the National Cancer Institute's Atlas of Cancer Mortality, 1950-1969, relies on a similar technique.[102]

[SS = (observed deaths-expected deaths)/sqrt(observed deaths)

where observed deaths = U.S. rate per 100,000 * population

* 20 / 100000]

However, these standardized rates cannot be compared across geography or time, as can the adjusted rates derived by the direct method, because each standardized calculation is based on a different age distribution to calculate the expected number of deaths.[103] [104] In other words, indirect standardization does not completely adjust for the differences in the composition of the populations under scrutiny.[105] This issue is one that is not well explored in the epidemiological literature. Indirect standardization is commonly used to account for the differences in the age distribution of the populations under study, but in fact it does not accomplish this goal.

---

[102]Mason T.J., et.al. *op.cit.*

[103]Mausner J.S., Bahn A.K. *Epidemiology: An Introductory Text.* Philadelphia: W.B. Saunders, 1974, p. 138.

Lilienfeld A.M., Lilienfeld D.E. *Foundations in Epidemiology, Second Edition.* New York: Oxford University Press, 1980, p.78-80.

[105]Fleiss J.L. *Statistical Methods for Rates and Proportions, Second Edition.* New York: Wiley and Sons, 1981, p. 244.

Because of the inability to compare across time or geographic area, truncated rates were used in the following analyses. Because the truncated rate are based on only the relevant portion of the population (women over age 35), the backround variation and differences in the distribution of older women are minimized, as illustrated in Table 2B.

4. ANALYSIS OF DATA BY STATE AND THE UNITED STATES

4.1. Time-trends

An examination of time-trends in age-adjusted mortality rates from ovarian cancer reveals a definite increase over time. This trend is shown in Table 3.

TABLE 3:														
Ovarian cancer: Age-adjusted death rates per 100,000; United States; White and non-white females; 1950-1978.														
	1950 -51	1952 -53	1954 -55	1956 -57	1958 -59	1960 -61	1962 -63	1964 -65	1966 -67	1968 -69	1970 -71	1972 -73	1974 -75	1976 -77	1978
WHITE	6.7	7.0	7.3	7.4	7.4	7.4	7.3	7.3	7.4	7.5	7.7	7.7	7.9	8.1	8.3
NON-WHITE	4.5	5.0	5.3	5.8	6.0	5.9	5.7	6.0	6.1	6.3	6.2	6.5	6.7	7.0	7.1

* Segi et.al. Mortality for selected sites in 24 countries. Tohoku Universiy, Japan, 1967.
** Herb Sauer Asso. Average Annual Age Adjusted Mortality. Columbia, MO.

Because age-adjustment often masks underlying patterns within specific age groups, the age specific rates for the years 1968-1978 were examined. As can be seen in Table 4, the overall age-adjusted mortality rates are being dominated by the increased rates in the oldest age group (65+) which rose 8% from 1968-1978 in white women, and 7% in non-white women. Mortality rates in the younger ages actually decreased over the same eleven year time period: 3% for both white and non-white women aged 25-44, 1% for white women aged 45-64, and 11% for non-white women.

This contrast in declining mortality rates in younger women versus increasing mortality rates in older women can be interpreted in one of two ways. One possible explanation is that the disease is being diagnosed at an earlier age allowing for effective therapeutic intervention. This would result

in a lower case fatality rate in younger women, but would not effect the rates of older women in whom the disease had already progressed beyond therapeutic control. Examination of published case-fatality rates by age for the three age groups did not support this hypothesis.[106] The second possible explanation is that older women experience an increased risk of developing the disease due to some cohort-specific risk factor or prior exposure. This hypothesis is explored in the following section.

---

[106]Axtell L.M., Asire A.J., Myers M.H. *Cancer Patient Survival*. Report No. 5, DHEW Pub. No. (NIH) 77-992, 1976.

TABLE 4:										
Ovarian cancer: Age-specific death rates per 100,00; United States females; 1968 - 1978.										
	1968	1969	1970	1971	1972	1973	1974	1975	1976	1977	1978
WHITE											
AGE											
25-44	6.5	6.1	5.8	5.3	5.3	5.2	5.5	4.7	4.6	3.8	3.9
45-64	43.1	44.4	44.1	43.8	43.8	42.5	43.8	42.6	42.2	42.0	41.8
65+	111.5	114.0	115.8	114.8	115.4	112.0	113.3	119.1	122.3	120.9	119.0
NON-WHITE											
AGE											
25-44	5.8	4.7	5.8	5.3	3.9	4.1	4.3	3.3	3.2	3.1	2.8
45-64	38.0	38.8	37.5	34.2	36.5	32.6	32.2	30.2	30.4	26.7	26.9
65+	76.4	68.6	66.1	82.0	83.3	82.8	80.4	68.8	81.8	83.9	83.7

4.2. Cohort analysis of mortality patterns

To determine if different generations of women experienced differing risk of ovarian cancer throughout their lifetimes, a cohort analysis was undertaken. To conduct the analysis, it was necessary to supplement the state level mortality data with data from the years 1953, 1958, 1963 by abstracting the death counts from the Vital Statistics of the United States. That such a cohort effect might underlie the observed mortality patterns is suggested by Valerie Beral's analysis of time trends in ovarian cancer mortality in England and Wales.[107] Beral demonstates that the cohort of women who born between the years 1900 - 1910 show an increase in mortality from ovarian cancer compared to those born before or after. She hypothesizes that this difference is due to the fact that these women, who were of child-

---

[107]Beral, *op.cit.*

bearing age during the Depression, had, on the average, fewer children than their predecessors or successors. It is known that fertility rates dropped markedly during the Depression in the United States (See Appendix; Item B). If one accepts the hypothesis of a protective effect from increased parity, one would expect to see an increase in ovarian cancer mortality among the cohort of women who were of childbearing age during the Depression. In fact, ovarian cancer mortality rates for all ages peak in the birth cohorts of 1899-1903 and 1904-1908, paralleling Beral's observations in England and Wales.

Oftentimes, epidemiologists rely on graphic representation to establish the existence of a cohort effect in a particular data set. A technique developed by Steve Selvin and Susan Sacks and described in their paper "A Method for Detecting a Cohort Exposure"[108] allows for a statistical assessment of the influence of birth cohort on mortality rates. Furthermore, this technique separates the effects of age and calander time from that of birth cohort to allow for assessment of their relative influence on disease rate by use of a linear model. In this method, disease rate is modeled as a function of some constant plus functions of age, calendar time and birth cohort. $r = u + a + b + c$, where i indicates the age categories, j indicates cross-sectional years, and (k) the birth cohort. The model produces an estimates rate by summing the estimated effects for each age group, each time period, and each birth cohort. Because each birth cohort is uniquely defined by age and calendar time, a value for the age/time interaction for each cohort, symbolized by $c(k)$, can be estimated from the model. A weighted least-squares method is used to derive the esti-

---

[108]Sacks S.T., Selvin S. A method for detecting a cohort exposure. *Environmental Research* 1981;25:167-177.

mates for the components of the linear model. An analysis of variance produces three F-statistics measuring the individual influences of age, calendar time, and birth cohort, and allows for tests of statistical significance. The lack of fit of the model to the observed data gives an estimate of the variation unexplained by age, time and cohort effects. The analysis was run using logarithmic transformation of the data to increase numeric stability as well as to produce an analysis that better fulfills the assumptions of an analysis of variance (i.e.- noramlly distributed data).

For this analysis, a total of 42 parameters were used to predict estimated rates: 15 age groups, 6 time periods, 20 birth cohorts, and 1 constant value. Given a total of 90 observations, it is clear that this model is grossly overparameterized, and results in a value for residual or unexplained variation that is deceptively low (Table 5).

TABLE 5		
	Whites	Non-whites
Regression	180.3428	123.4277
Residual	.6262	6.6615
Total	180.9614	130.0904

In other words, the predicted data fit the observed data almost perfectly. Because of this, it is not possible to rely on formal tests of statistical significance to assess the contribution of age, time period, and birth cohort. However, the predictive value of the model after removing the effect of one of these three can be compared, generating a quantitative assessment of their relative importance in predicting disease rate over time. Table 6 presents the ratios of the predictive power of the model with the effect of cohort, period and age removed. The ratios indicate the relative influence of these three effects in the overall prediction of mortality.

TABLE 6		
Whites	Sum of squares	Ratios
Cohort	0.241	2.5
Period	0.096	1.0
Age	1.655	17.0
Non-whites		
Cohort	0.866	2.3
Period	0.379	1.0
Age	1.574	4.2

As can be seen, the effect of age is expectedly the most powerful of the three. It is seventeen times more influential than the effect of time period in predicting mortality rates among whites, and four times as influential among non-whites. In both cases, the effect of birth cohort is intermediate to the others, indicating to the existence of a cohort effect in ovarian cancer mortality. This analysis shows that year of birth is more important in predicting ovarian cancer mortality than is calendar time. One possible explanation for this finding, as mentioned above, is the confounding effect of parity. Nulliparous women have an increased risk of ovarian cancer, and the risk decreases progressively with increasing parity, as discussed in Section 2.2. This relationship is quite strong: women with no children have a four times greater risk of ovarian cancer than women with four or more children.[109] [110] When the ovarian cancer mortality rate is plotted against the corresponding fertility rate, a strong negative association is seen ($r = -0.78$; $p < 0.01$). This pattern was also found by Valerie Beral in her analysis of data from 1861 to 1931 in England and Wales, and her analysis revealed that the more than two-fold increase in the age- adjusted mortality rate from ovarian cancer was explained largely by changes in th average completed family size.[111] This

[109]Newhouse et.al.,

[110]Casagrande et.al., op.cit.

[111]Beral V., Fraser P., Chilvers C. Does pregnancy protect against ovarian cancer? Lancet 1978; May 20: 1083-1087.

ecological association will be further analyzed in Section 5.1 - 5.3 using county level data.

This technique allows the graphic display of the magnitude of the cohort, time, or age effects as done by Ernster, Selvin and Winkelstein in their analysis of prostatic cancer mortality among non-whites.[112] Ovarian cancer rates among white and non-white females with the birth cohort, time period and age effects statistically removed are shown below (Figures 2- 9), and again, the cohort effect is apparent.

---

[112]Ernster V.L., Selvin S., Winkelstein W. Cohort mortality for prostatic cancer among United States non-whites. *Science* 1979; 200:1165-1166.

FIGURE 2: Estimated ovarian cancer mortality per 100,000;
   United States white females aged 35 and older;
      Including the effects of age, period (calendar time),
      and birth cohort; 1953, 1958, 1963, 1968, 1973, 1978.

FIGURE 3: Estimated ovarian cancer mortality per 100,000;
   United States white females aged 35 and older;
      Including the effects of age and period (calendar time),
      and removing the effect of birth cohort;
      1953, 1958, 1963, 1968, 1973, 1978.

FIGURE 4: Estimated ovarian cancer mortality per 100,000;
    United States white females aged 35 and older;
        Including the effects of age and birth cohort,
        and removing the effect of period (calendar time);
        1953, 1958, 1963, 1968, 1973, 1978.

FIGURE 5: Estimated ovarian cancer mortality per 100,000;
    United States white females aged 35 and older;
        Including the effects of period (calendar time)
        and birth cohort, and removeing the effect of age;
        1953, 1958, 1963, 1968, 1973, 1978.

FIGURE 6: Estimated ovarian cancer mortality per 100,000;
   United States non-white females aged 35 and older;
      Including the effects of age, period (calendar time),
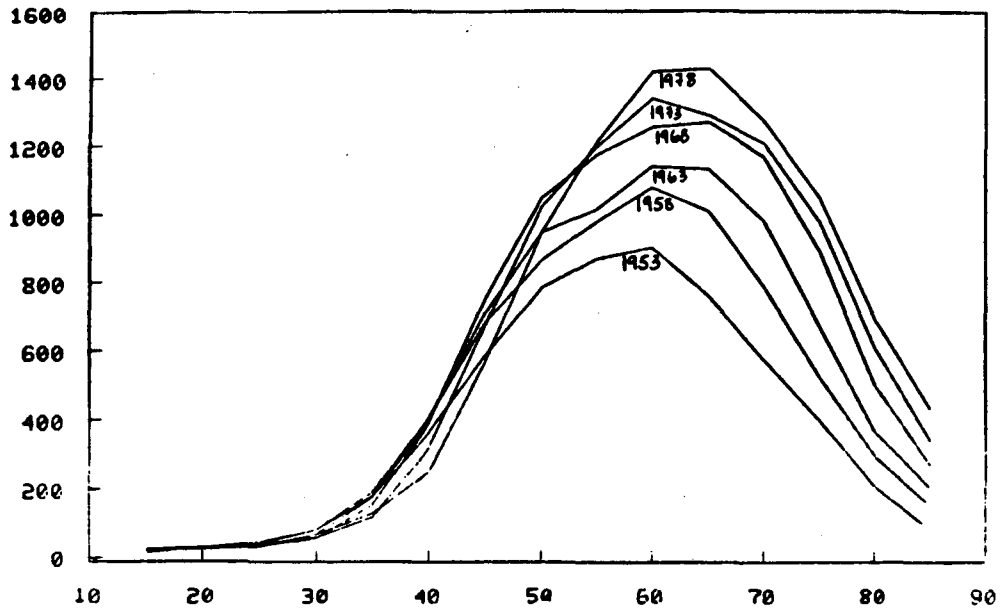   and birth cohort; 1953, 1958, 1963, 1968, 1973, 1978.
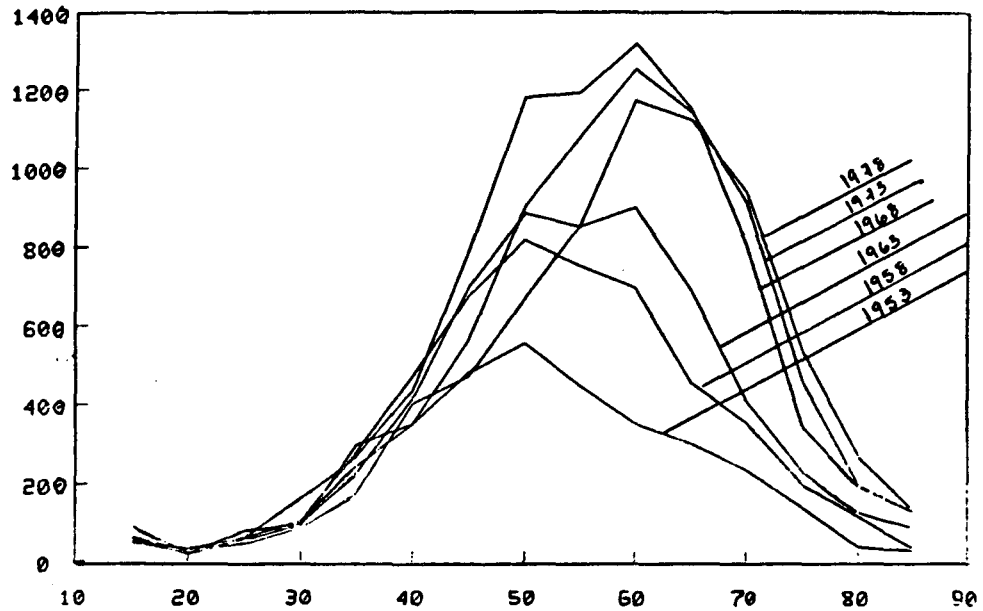
FIGURE 7: Estimated ovarian cancer mortality per 100,000;
   United States non-white females aged 35 and older;
      Including the effects of age and period (calendar time),
      and removing the effect of birth cohort;
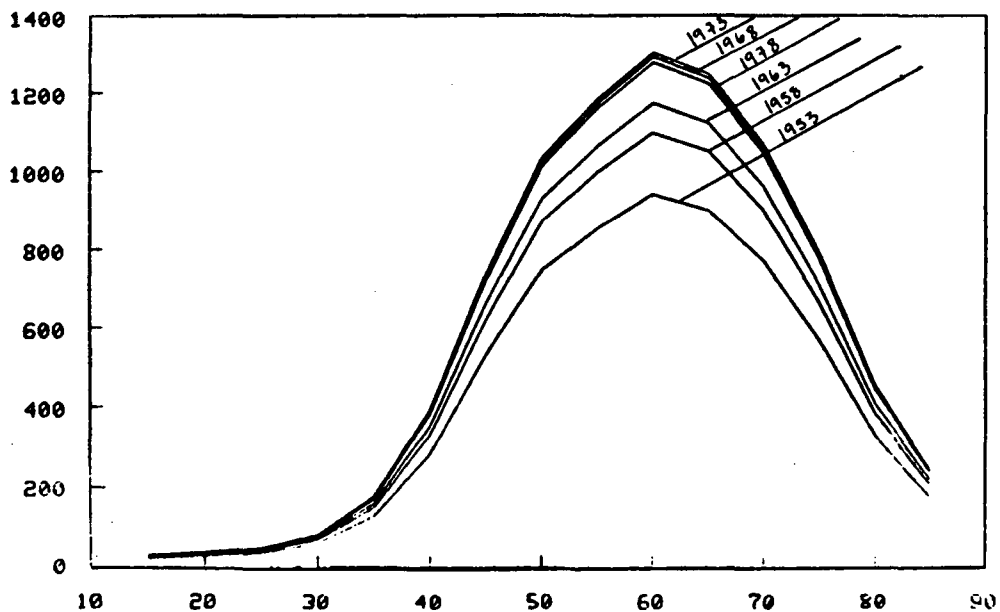   1953, 1958, 1963, 1968, 1973, 1978.
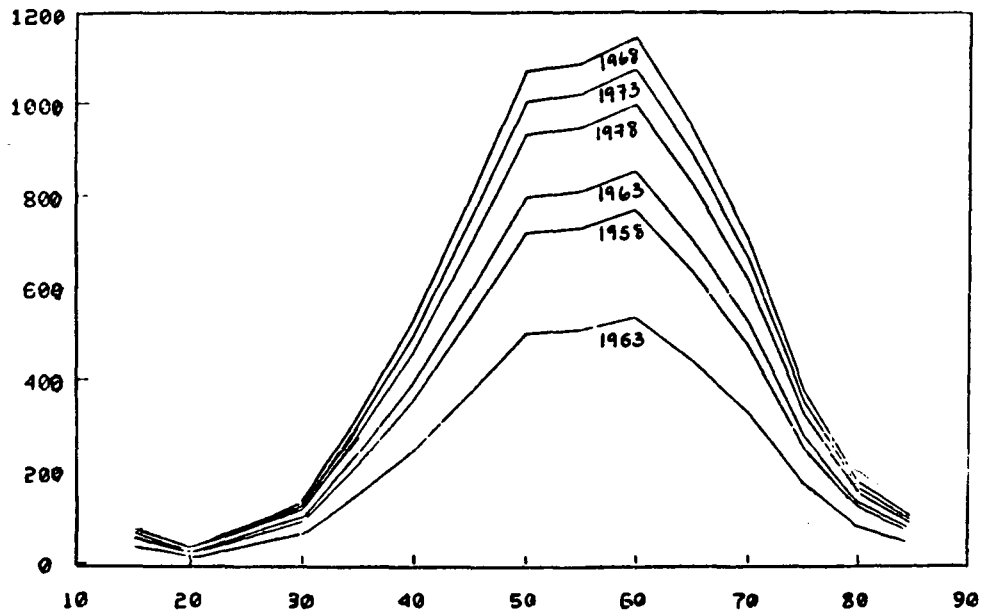
FIGURE 8: Estimated ovarian cancer mortality per 100,000;
United States non-white females aged 35 and older;
Including the effects of age and birth cohort,
and removing the effect of period (calendar time);
1953, 1958, 1963, 1968, 1973, 1978.
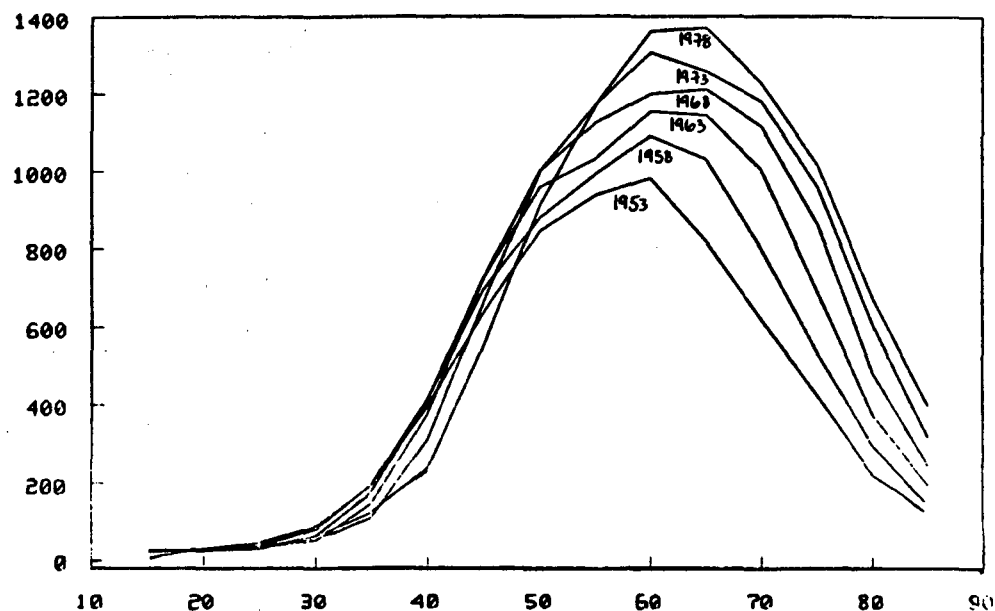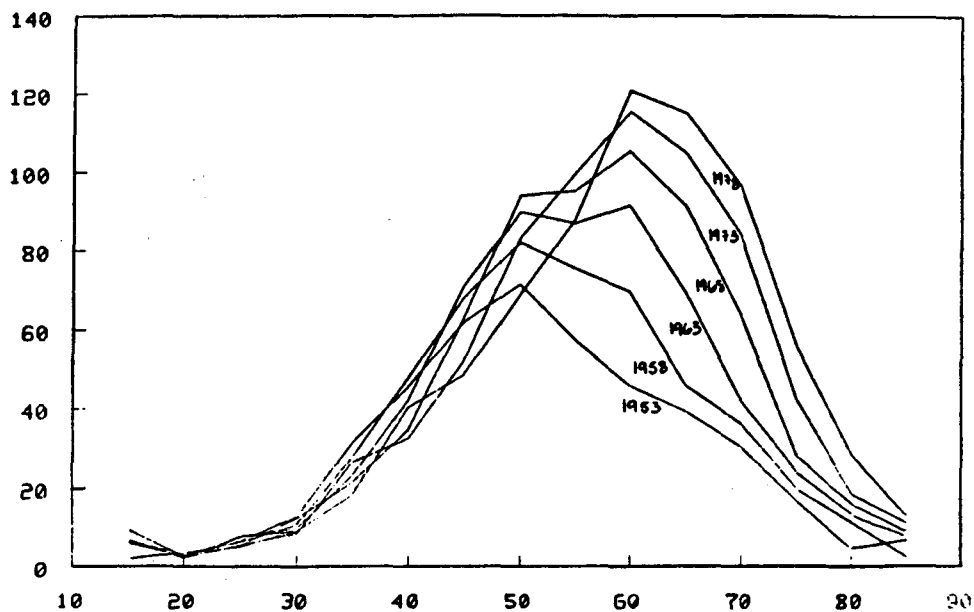

FIGURE 9: Estimated ovarian cancer mortality per 100,000;
United States non-white females aged 35 and older;
Including the effects of period (calendar time)
and birth cohort, and removeing the effect of age;
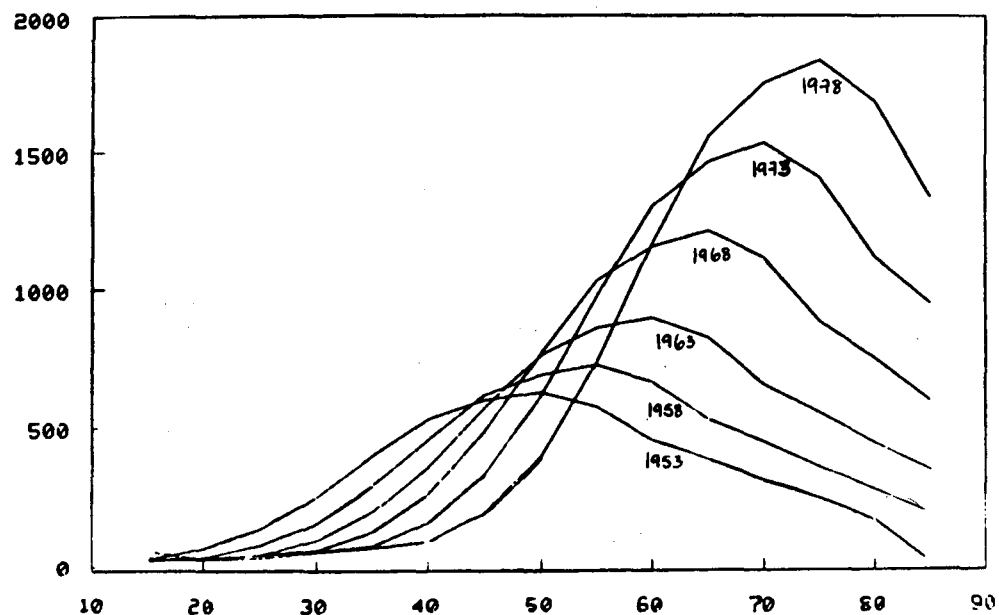1953, 1958, 1963, 1968, 1973, 1978.
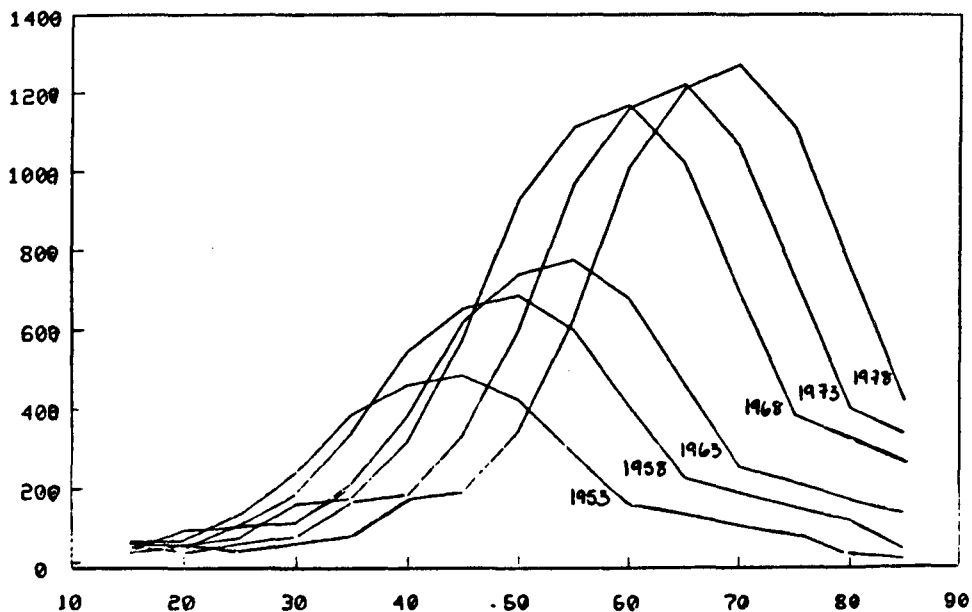
A clear graphic representation of the influence of the cohort effect on ovarian cancer mortality can be seen in Figure 10, using a technique described by Ernster et.al..[113] The age-adjusted ovarian cancer mortality rates among white females with the cohort effect statistically removed are plotted along with the observed rates. The modeled rates are very similear to the observed rates until 1968 when they begin to decline. The divergence between the observed and the modeled rates is directly attributable to the cohort effect.

FIGURE 10: Annual age-adjusted ovarian cancer mortality rates per 100,000 United States white females aged 35 and older; 1953-1978.

---
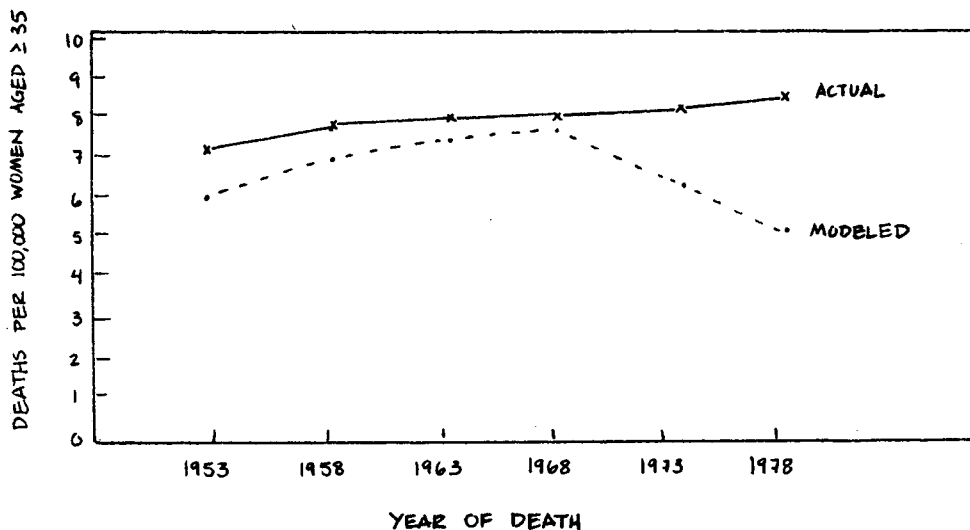
[113]Ernster V.L., Selvin S., Winkelstein W., *op.cit.*

4.3. Geographic patterns

The state level map of the average annual mortality rates from ovarian cancer 1968 - 1978 (Map 1) reveals a concentration of high rate states in the North-central and North-eastern United States. This pattern is repeated in the map of percent change in ovarian cancer mortality from 1968 to 1978 (Map 2). The high rates for the eleven year period are those states in which ovarian cancer mortality rose more than 8.4 per 100,000 between 1968 and 1978, specifically: California, North Dakota, Nebraska, Iowa, Missouri, Illinois, Wisconsin, New York, New Jersey, Delaware, New Hampshire, and Massachusetts. This pattern will be analyzed in greater detail using county data in Section 5.4.

MAP 1: Ovarian cancer mortality per 100,000 women, white and non-white, age >35; Average annual rate, 1968 - 1978.

MAP 2: Ovarian cancer mortality per 100,000 women, white and non-white, age >35; Percent change, 1968 - 1978.
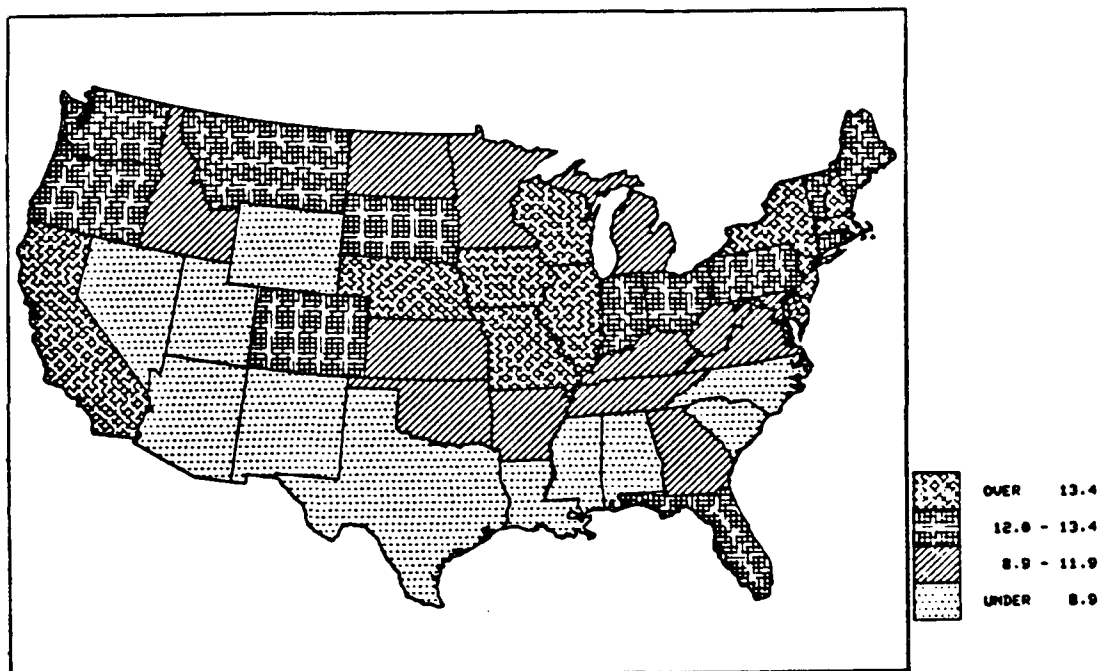
Legend:
- OVER 8.42
- 7.31 - 8.42
- 6.58 - 7.30
- UNDER 6.58
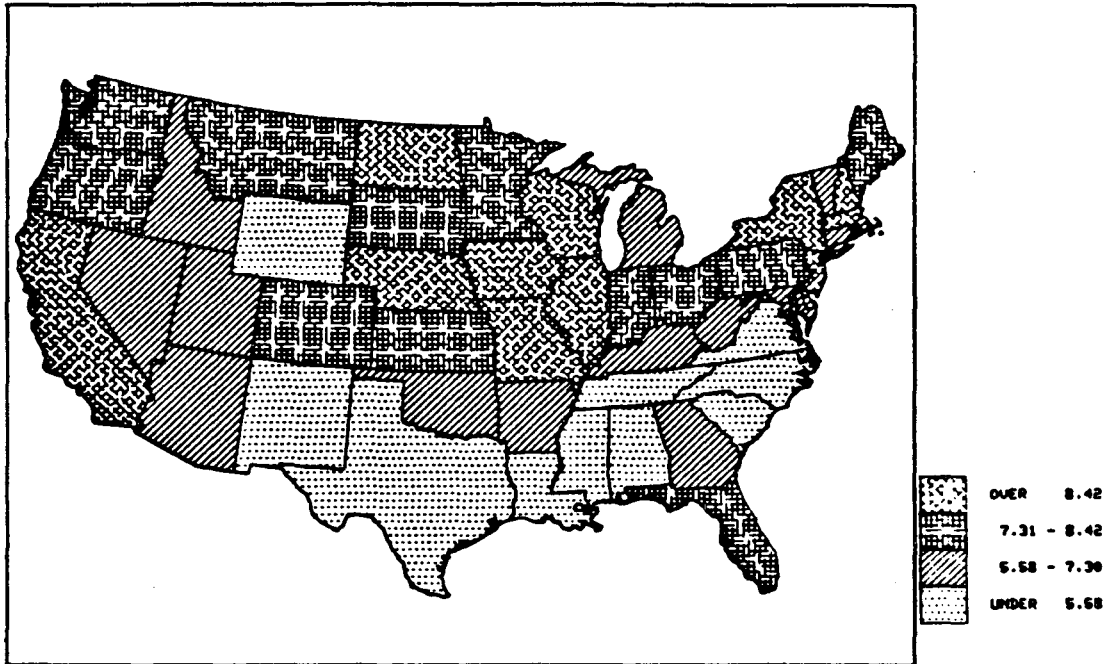
5. ANALYSIS OF DATA BY COUNTY

5.1. Ecological regression analysis

To further investigate the patterns and correlates of ovarian cancer, county mortality data for the eleven year period 1968 - 1978 were analyzed by socio-demographic variables of interest. Ecological regression analysis is useful as a descriptive tool summarize the linear dependence of an outcome on one or several predictor variables while controlling for the confounding effects of other factors. The results of the analysis indicate the variability in the outcome accounted for by the influences of the predictors.

Choosing the variables to include in the regression analysis is often rather arbitrary. The inclusion of variables in a regression analysis can sharpen a contrast, or it can dull it, especially if one allows a stepwise regression procedure to determine which of the many predictor variables is important. The gain or loss in precision will depend on how strongly the other variables influence the response being studied.[114] In the initial regression analysis, the dependent variable, ovarian cancer mortality rate, 1970, white and non-white females, age 35 and over, was analyzed using four independent variables as predictors. All four had evidence of an association with ovarian cancer in the literature, and each had a Pearson's coefficient of at least + or - 0.30 when correlated with ovarian cancer mortality. The variables analyzed were:

x1. Percent urbanization, 1970;
x2. Birthrate per 1000 population, 1970;
x3. Percent families earning <$15,000 per year, 1970;
x4. Number of physicians per 100,000 residents.
This model explains 46% of the variation in ovarian cancer mortality ($F = 3.43$; $P < 0.05$). The contribution of the specific variables can be examined through the regression coefficients. The regression coefficient Bi represents

---

[114]Hanley J.A. Appropriate uses of multivariate analysis. *Ann. Rev. Public Health* 1983; 4:155-180.

the change in the outcome Y with a one unit change in the predictor variable X_i, with all of the other predictors held constant. The coefficients are measurement dependent, however, and the fact that the independent variables are measured in different units necessitates the use of standardized coefficients to compare their relative influence on the outcome. The coefficients and their statistical significance (t-test) are presented below.

Variable	Regression Coefficient	Statistical Significance
Percent urbanization, 1970	0.43	0.02
Birthrate per 1000 population, 1970	-0.33	0.03
Percent families earning <$15,000	-0.09	0.53
Number of physicians per 100,000	0.12	0.52

In a series of "F-to-remove" tests, two variables were found to have a significant effect on ovarian cancer mortality in 1970: percent urbanization and birthrate per 1000 population ($F = 2.60$; $p < 0.05$). The other variables were not found to be significant predictors.

5.1.1. Choice of predictor variables

The fact that cancers have such a long latency period presents unique problems in epidemiological research. In an ecological analysis such as this, it is debatable whether the independent variables of interest should be measured at that point in time when their purported effect was likely to occur, or, as is often done, at the same point in time as the dependent variable (whether it be incidence or mortality data). In other words, would an investigation of the effects of fertility on ovarian cancer mortality be most meaningful if it examined the birthrate of 1940 and ovarian cancer mortality in 1970, assuming a 30 year latency for development of the cancer? There are assumptions underlying the use of predictor variables measured con-

currently with mortality or incidence rates, as well as the use of those meas-
ured to account for a latency in disease development. In using concurrent
measurements, one assumes that the factors of interest have remained rela-
tively stable over time. If however, there have been marked fluctuations in
the factors, it could be argued that one should consider the 20-30 year
latency factor in cancer development and measure the factors at the time of
cancer induction. In using the measurements which account for the latency
in the disease, one assumes that the populations in the counties under scru-
tiny have remained relatively stable, i.e.- there has not been significant
in/out migration. These assumptions must be considered and weighed with
respect to each variable.

The predictor variables of interest were expanded to include:
urbanization, birthrate, socioeconomic status, physician/population ratio
and employment of women. Concern over the biological timing of events led
to the use of 1940 measurements for certain of the predictor variables. After
examining fertility patterns from 1940 - 1970 (see Appendix item B), it was
decided to use 1940 birthrate statistics in the regression analysis. Fertility
rates fluctuated quite dramatically during this 30 year period, and the
change was relatively consistent across the United States. This geographic
stability diminishes the concern over the assumption of population stability
mentioned above, because although the women experiencing ovarian cancer
in a particular county may not be the same women whose contributed to the
birthrate of 1940, their "exposure" to the suspected etiologic agent (i.e. -
years of ovulation) was similar. By the same reasoning, 1940 female employ-
ment statistics were used in the equation. This variable, percent of women
employed in 1940, might influence the timing and number of children a
woman chose to have. It is expected that counties with high levels of female

employment will show relatively lower birthrates, and ultimately, this should be reflected in higher levels of ovarian cancer mortality.

The 1970 measurements were used for the other variables because their purported effect is not on disease etiology but rather on disease detection and diagnosis. The positive associations between ovarian cancer incidence and mortality and higher socioeconomic status or incresed urbanization is probably due to some extent to differntial access to medical care. In other words, women of higher SES or those living in more urbanized areas are more likely to have their cancer detected and diagnosed compared to women of lower SES or those living in more rural areas. Because ovarian cancer is a disease which is difficult to diagnose, its incidence and mortality may be under-represented in areas where there is poor access to medical care, i.e.- more rural areas, areas of lower SES, and areas with fewer physicians. This ascertainment bias could serve to artificially inflate the association between ovarian cancer mortality and cwcertain predictor variables.

In these data, the independent variables measuring percent urbanization, socio-economic status, and physician/population ratio all contribute to a woman's access to medical attention and care. Principal component analysis, a statistical technique which reduces multivariate measurements to one or a few summary numbers (canonical variables), was used to estimate the best linear combination of these three highly correlated variables and to give one summary measure of access to medical attention and diagnosis. The analysis resulted in the formation of a summary variable which combined measures of urbanization, income and physician/populaton ratio, and accounted for 71% of the variation. This summary variable was used in all subsequent analysis.

5.1.2. Results of regression analysis using 1940 and 1970 predictors

The regression analysis was run using the dependent variable ovarian cancer mortality per 100,000, 1970, and the following independent variables:

x1. A summary variable designed to measure access to medical attention and diagnosis and comprised of measures of income, urbanization and medical manpower;
x2. Birthrate per 1,000 residents in 1940;
x3. Percent of women aged >14 employed in the laborforce in 1940.
The overall predictive value of the equation was improved by the use of a combination of 1940 and 1970 predictors (R = 0.46 vs. R = 0.62; F = 3.94, p < 0.05), a finding which is intuitively pleasing and supports the idea of considering latency in ecological studies. The contributions of the specific variables are presented below.

Variable	Regression Coefficient	Statistical Significance
Access to medical care	0.34	0.04
Birthrate per 1000 population, 1940	-0.69	0.02
Percent of women employed, 1940	-0.03	0.64

As can be seen, the first two variables listed contribute the most to the understanding of ovarian cancer mortality. In a series of "F-to-remove" tests, these two variables were found to be significant predictors (F = 3.99, p <0.05), while percent of women employed, 1940 was not.

5.2. Alternative to ecological regression analysis

5.2.1. Introduction to the Organized Categories of Equal Risk (OCER) technique

The method of ecological regression analysis has been much maligned in the epidemiological literature on the basis of its numerous sta-

tistical and interpretational pitfalls, including the ecological fallacy.[115] In spite of the numerous critiques of use of the ecological approach to study the relationship between disease patterns and certain epidemiologic factors, few alternatives have been developed. One approach that does not depend on statistical models, yet produces easily interpreted results that closely parallel those of ecological regression analysis, has been developed to analyze nationwide, geographically based data. The procedure, referred to as the Organized Categories of Equal Risk (OCER) technique, has been described elsewhere,[116] but a brief description of the technique is presented here.

Suppose one were interested in the relationship between urbanization and mortality from a particular disease. A data file, organized by geographic area, e.g.- counties in the United States, is compiled. For each area, the following data values are included: 1. independent variable , e.g.- percent of urbanization; 2. dependent variable , e.g - number of deaths due to ovarian cancer; 3. estimate of person years at risk, e.g - number of women aged > or = 35. The file is ordered from high to low on the basis of the independent variable value, and the number of deaths are then accumulated into a series of groups with equal numbers of person years of risk. The file is now composed of a series of equal risk categories. If an ordering is induced in the dependent variable by the ordering of the independent variable, it is inferred that an association exists between the two variables. Under the null hypothesis of no association between degree of urbanization and mortality from ovarian cancer, the expected number of deaths in each cell is equal to the overall expected mean value, and ithe number of deaths should follow a

[115]Robinson W.S. Ecologocal correlations and the behavior of individuals. *Amer Soc Rev* 1950; 15:351-357.

[116]Selvin S., Merrill D., Sacks S. An alternative to ecological regression analysis of mortality rates. *Amer J Epidemiol* 1982;115:617-623.

Poisson distribution. Therefore, the null hypothesis of no association between the predictor and the outcome is easily tested using the chi-square test of variance, contrasting the observed variation in the number of deaths with the expected variation given a Poisson distribution. The ratio of these estimated variances should be approximately equal to 1.0 when no association exists. A nice property of the chi-square test applied to equal risk categories, pointed out by the authors of the technique, is that it is a conservative test.[117] In other words, if one finds evidence of an association using aggregated data, it is more likely that an association exists in the ungrouped data. This is not the case when one uses regression coefficients to test for an association between two ecologic variables.

This technique was utilized to assess the influence of a series of predictor variables on county ovarian cancer mortality rates. The results are then compared and contrasted with the results of an ecological regression procedure.

5.2.2. Results of OCER analysis

A file was developed using 1970 county level mortality and population data for white and non-white females, and data on the following predictor variables:

x1. Percent families earning <$15,000 per year, 1970;
x2. Percent of families below the poverty line, 1970;
x3. Percent of urbanization in the county, 1970
x4. Number of physicians per 100,000 residents, 1970;
x5. Birthrate per 1,000 population, 1940;
x6. Percent of woman aged <14 employed in the laborforce;
x7. A summary variable designed to measure access to medical attention and diagnosis and comprised of measures of income, urbanization and medical manpower.

The file was then divided into 50 equal risk categories on the basis of one of these variables (known as the c-variable) and the null hypothesis of

---

[117]*Ibid.*, p.320.

no association between this predictor variable of interest and ovarian cancer mortality (the a-variable) was tested using a chi-square statistic, while controlling for the effect of the potential confounder (known as the b-variable). That is, an association between variables a and c was investigated while controlling for the potential confounder, variable b. The OCER analysis also yields a measure analogous to the squared multiple correlation coefficient commonly used in regression analysis and measuring the percentage of the total variation in the outcome variable which is accounted for by the predictor. This value was used to calculate a ratio similar to those used in the "F-to-remove" tests cwcommonly used in multiple regression analysis. The percentage of the total variation in ovarian cancer mortality that is explained by: 1) the combination of the b and c variables, and, 2) the c-variable while controlling for the b-variable was compared to the maximum variance (or predictability) when the number of eaths servedd as both the predictor and the outcome.[118] The latter ratio is comparable to an "F-to-remove" ratio in which the predictive power of a model with variable x1 removed is compared to the predictive power of the model incluing the variable in order to assess its influence on the outcome.

OCER analyses were used to investigate the following questions:

1. Does the extent of urbanization in a particular county have an independent effect on ovarian cancer mortality apart from a)the effects of income, or b)number of physicians?
2. Does income have an independent effect on ovarian cancer mortality apart from a)birthrate or b)number of physiscians?
3. Does the percentage of individuals below the poverty level in a county have an independent effect on ovarian cancer mortality apart from urbanization?
4. Does birthrate have an independent effect on ovarian cancer mortality apart from access to medical care?
5. Does access to medical care have an independent effect on ovarian cancer mortality apart from a)birthrate and b)poverty status?
6. Does birthrate have an independent effect on ovarian cancer mortality apart

[118]Selvin S., Merrill D., Sacks S. An alternative to ecological regression analysis of mortality rates. *Amer J Epidemiol* 115:617-23, 1982, p. 620.

from percent of women in the laborforce?
The results of the OCER analysis are presented in Table 7.

TABLE 7:							
RESULTS OF CO-CHOP ANALYSES							
			RATIO OF PARTIAL R TO TOTAL R				
VAR B	VAR C	N	DEATH BY B	DEATH BY C	DEATH BY C.B	CHI-SQUARE	P-VALUE
INCOME	URBAN	2906	46.5%	72.4%	2.0%	0.28	N.S.
DOCSRT	URBAN	2906	24.3%	72.4%	12.9%	0.13	N.S.
BRTHRT	INCOME	2906	12.4%	55.4%	20.1%	6.7	<.001
DOCSRT	INCOME	2906	24.4%	55.4%	17.1%	5.5	<.001
URBAN	POVERTY	2904	72.4%	20.4%	10.3%	3.0	N.S.
ACCESS	BRTHRT	2904	74.5%	12.4%	10.2%	6.1	<.001
BRTHRT	WOMLAB	2906	12.4%	39.1%	11.7%	2.2	N.S.
BRTHRT	ACCESS	2886	11.4%	76.2%	72.2%	57.3	<.001
POVERTY	ACCESS	2886	20.6%	76.2%	66.6%	66.2	<.001

The figures in columns 4-6 compare the predictive power of variable
b (column 4), variable c (column 5), and variable c controlling for variable b
(column 6), to the maximum predictability, and indicate those variables that
are important independent predictors of ovarian cancer mortality. The
results are then tested using a chi-square statistic to determine if the
observed variation in ovarian cancer deaths is significantly different from the
expected variation, assuming a Poisson distribution. These analyses revealed
that percent of families earning <$15,000 in 1970, birthrate in 1940, and
access to medical care are all significant predictors of ovarian cancer mor-
tality.

5.3. Comparison of ecological regression and OCER analyses

The results from the ecological regression analysis and the OCER
analysis are very similar. The OCER analyses showed percent of families
earning < $15,000 in 1970, birthrate in 1940, and access to medical care to
be important independent predictors of ovarian cancer mortality. The chi-
square values for the OCER analyses for these variables, while controlling for
a third variable, were all highly significant, indicating a significant divergence
from the expected Poisson distribution of deaths under the null hypothesis of
no association. In the initial regression analysis, the "F-to-remove" tests
indicated that the most important predictors of ovarian cancer mortality

were percent urbanization in 1970 and birthrate in 1970. In the second regression analysis, using a combination of 1940 and 1970 variables, the most important predictors were found to be access to medical care and birthrate per 1,000 population.

The key problem in any ecological regression analysis is the lack of information about the joint distribution of the study factor and the disease within each group (i.e. - unit of analysis). This can lead to substantial bias and can either artificially inflate or deflate the ecological association. This bias, commonly referred to as the ecological fallacy, can be partitioned into two components: 1) aggregation bias - due to the grouping of individuals; and 2) specification bias - due to the confounding effects of the group itself.[119] In this latter component, either some extraneous risk factor is differentially distributed by group, or some property of the group itself affects the out-come.[120] It is quite likely that such a bias affected the findings in the regression analyses, but there is no means for testing for or measuring this bias. One advantage of the OCER technique is the fact that it is not subject to either aggregation or specification bias. However, it is also a less statistically sophisticated measure of association than the correlation coefficient, and may overestimate the association.[121]

5.4. Geographic patterns

5.4.1. Introduction

A technique similar to OCER was employed by Blair et.al. in their

---

[119]Morgenstern H. Uses of ecological analysis in epidemiological research. *AJPH* 1982;72:1336-1344.

[120] *Ibid., p.1339.*

[121]Selvin; Personal communication.

investigation of geographic patterns of leukemia in the United States.[122] Age-adjusted mortality rates for leukemia were correlated by race and sex with demographic, industrial and agricultural data for 3056 U.S. counties. The county mortality rates were also related to demographic, industrial and agricultural variables using a weighted multiple regression model with the weights directly proportional to the square root of the counties' total person years at risk, that is inversely proportional to the standard error of the estimated mortality rates. The investigators found that certain geographic patterns persisted even after adjusting for ecological variables (through stratification along regional urbanization and socioeconomic lines). The fact that the regional variation in leukemia mortality persisted after adjusting for demographic differences suggests that there are geographically related etiologic factors. Inspired by this result, a similar analysis was conducted of geographic variation using the OCER technique. This procedure is described below.

5.4.2. OCER analysis of 1970 ovarian cancer mortality

The advent of computer mapping techniques has led to an increase in geographic analyses of mortality and/or incidence patterns. Such maps are useful tools in both the development and the testing of etiological hypotheses, and analyses of county mortality data by certain demographic and environmental characterisitcs have proliferated. An analysis of geographic variation in ovarian cancer mortality at the county level using the OCER technique has distinct advantages over the traditional methods of computerized geographic analyses. This technique divides the country into a series of equal risk categories with respect to latitude or longitude, thereby

---

[122]Blair A., Fraumeni, J.F., Mason T.J. Geographic patterns of leukemia in the United States. *J Chron Dis* 1980;33:251-259.

removing the problem of the larger counties (which are concentrated in the Western United States) having a disproportionate visual impact on a map of the entire United States.[123] This problem is evident in the Atlas of Cancer Mortality for United States Counties, 1950 - 1969, published by the National Cancer Institute, and in the county level maps generated by the SEEDIS program that follow. In these maps, there is no allowance made for the variation in population size among the counties. Therefore, a chance increase in the number of deaths due to disease X in a relatively small county may appear as a significant deviation from the U.S. mean; this county would appear as a high rate county on a computer generated map of disease X mortality. This is particularly a problem in investigations of rare diseases, a fact that was pointed out by the authors of the Atlas.

The OCER technique uses latitude or longitude (or some combination thereof) as a predictor variable and sorts the number of deaths from a particular disease into a series of geographic strips having equal populations at risk. One can then test for geographic patterns by contrasting the observed variation in mortality with that expected under a Poisson distribution. This was done using 1970 ovarian cancer deaths and three geographic measurements: latitude, longitude, and a lat-long index. This index, a two-dimensinal parabolic function, is suggested by Selvin et.al. in their initial paper on the OCER technique.[124] In all three analyses, the effect of urbanization was controlled by statistical means. The research questions to be answered were:

1. Is there evidence of geographic variation in ovarian cancer mortality from west to east (latitudinally)?
2. Is there evidence of geographic variation in ovarian cancer mortality from north to south (longitudinally)?

[123]Selvin S., Merrill D., Sacks S., *op.cit.*, p 621.

[124]Selvin S, Merrill D., Sacks S., *op.cit.*, p.622.

3. Is there evidence that there are lower rates of ovarian cancer in the central states as compared to the border states?

The results of the three analyses to test these hypotheses can be

seen in Table 8.

TABLE 8:		
Geographic analysis of ovarian cancer mortality, 1970 United States females, White and Non-white.		
Predictor	Chi-square=	P value=
Latitude	0.03	0.86
Longitude	0.04	0.55
Latlong index	5.60	0.01

There is no evidence of either latitudinal or longitudinal geographic variation, but it appears that the two dimensional parabolic function described by the lat-long index does describe the pattern of ovarian cancer mortality. This is consistent with the findings of the state level map of average annual mortality rates, 1968 - 1978. However, it is not at all consistent with the results of the county level mortality maps, as described in the following section.

5.4.3. Ovarian cancer mortality for U.S. counties 1968 -1978

County level maps were generated to display the mortality rates due to ovarian cancer (truncated crude rates, as described in Section 3.2.3.3). A map of the 1970 rates (Map 3) was generated to insure comparability with the previously described OCER geographic analyses. Maps of the 1968 - 1978 average annual rate (Map 4) and the percent change from 1968 to 1978 (Map 5) were also generated. The maps do not reveal any consistent geographic pattern. The percent of counties falling in the highest mortality rate category (defined as those counties with an average annual rate >18.9% for the years 1968-1978) within each of the ten federal regions of the United States are shown in Table 9. The high rate counties appear to be randomly scattered throughout the United States, in contrast to the results reported by NCI indicating concentrations of high rate counties in the rural North and low rate counties in the South.

TABLE 9:	
Percent of counties in high mortality rate category in each federal region; United States females, white and non-white; age > 35.	
Federal region	Percent of counties falling in high-rate category
I. New England	13.0
II. New York-New Jersey	15.5
III. Middle Atalntic	11.0
IV. Southeast	13.4
V. Great Lakes	14.2
VI. Southcentral	16.6
VII. Central	15.2
VIII. Mountain	9.2
IX. West	11.3
X. Northwest	8.4

States inculded in Federal regions:
I. Maine, New Hampshire, Vermont, Massachusetts, Rhode Island, Connecticut
II. New York, New Jersey
III. Delaware, District of Colombia, Pennsylvania, Maryland, Virginia, West
    Virginia
IV. Alabama, Florida, Georgia, Kentucky, Mississippi, North Carolina, South
    Carolina, Tennessee
V. Illinois, Indiana, Michigan, Minnesota, Ohio, Wisconsin
VI. Arkansas, Louisiana, New Mexico, Oklahoma, Texas
VII. Iowa, Kansas, Missouri, Nebraska
VIII. Colorado, Montana, North Dakota, South Dakota, Utah, Wyoming
IX. Arizona, California, Hawaii, Nevada
X. Alaska, Idaho, Oregon Washington

MAP 3: Ovarian cancer mortality per 100,000 women,
White and non-white, age >35; 1970.
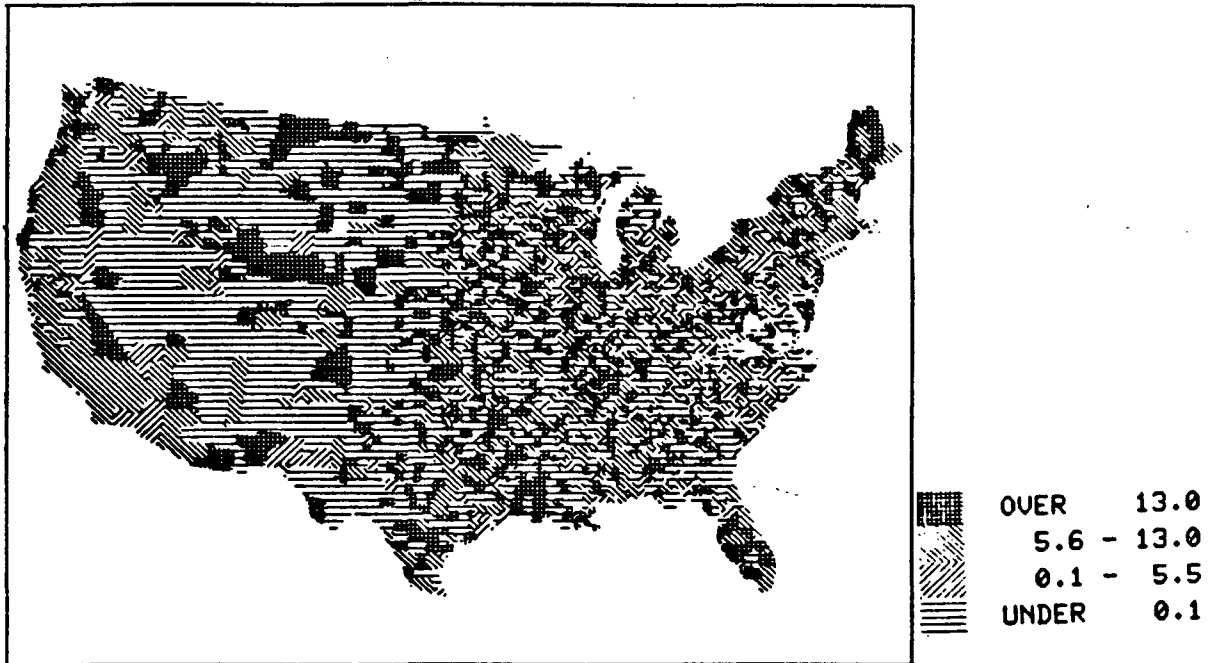
OVER   13.0
5.6 - 13.0
0.1 - 5.5
UNDER   0.1

MAP 4: Ovarian cancer mortality per 100,000 women,
White and non-white, age >35; Average annual rate, 1968 - 1978.

OVER   18.9
11.2 - 18.9
1.8 - 11.1
UNDER   1.8

MAP 5: Ovarian cancer mortality per 100,000 women,
       White and non-white, age >35; Percent change, 1968 - 1978.

OVER	13.3
6.7 -	13.3
0.0 -	6.6
UNDER	0.0
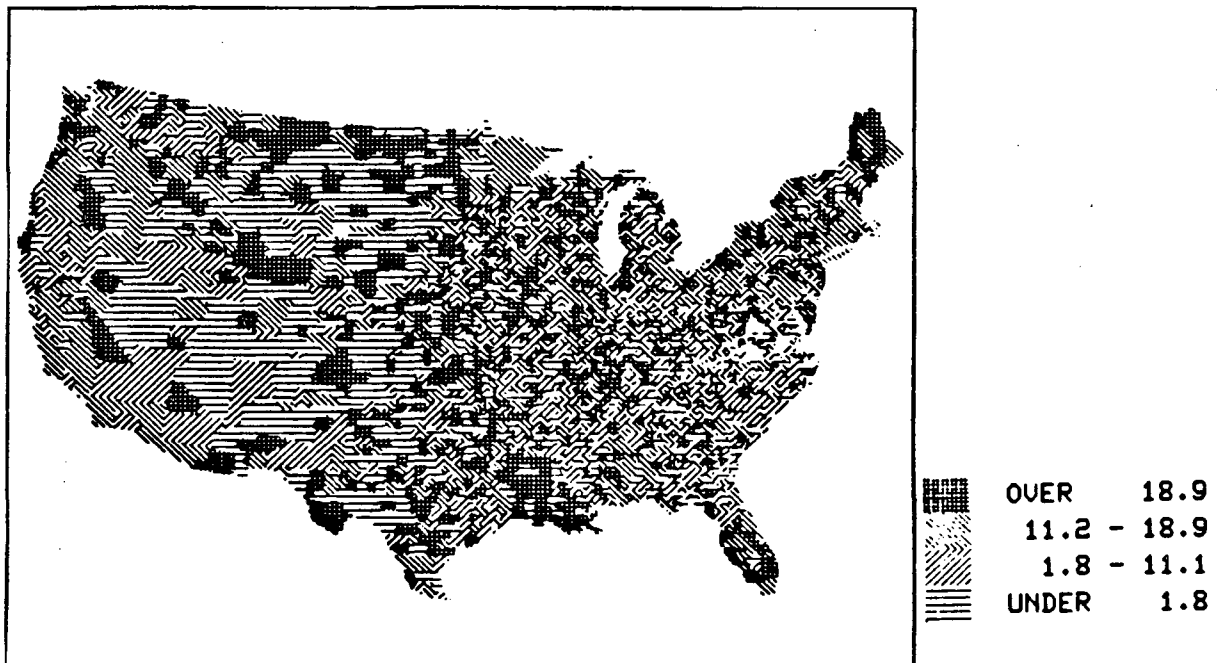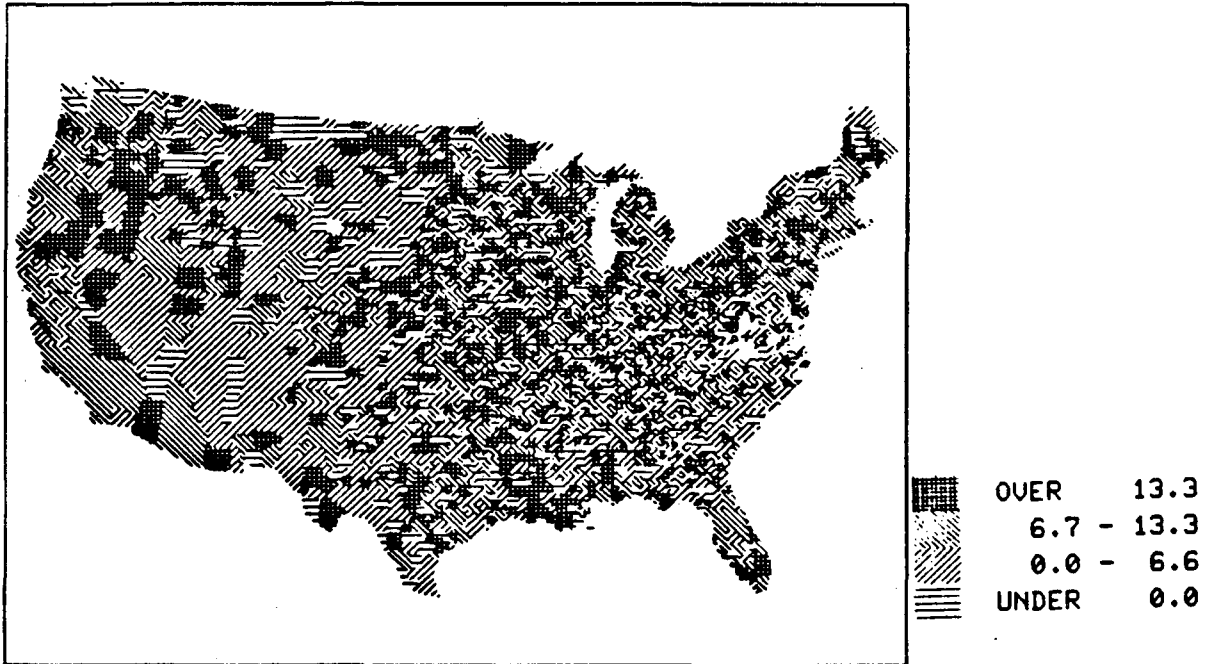
5.4.4. Discussion of geographic analyses

It is important to discuss the differences in the mortality patterns found by the two approaches. The traditional computer mapping technique, e.g.: as used in the National Cancer Institute's Atlas of Cancer Mortality: 1950 - 1969 and in the above maps of county level mortality from ovarian cancer, 1968 - 1978, does not account for the variation in county size, either in terms of geography nor in terms of population. Thus, a county with a very large surface area has a disproportionate visual impact on a map of the entire United States. Furthermore, a small county with a very few number of deaths and a relatively unstable mortality rate may appear as a high risk county due to a random fluctuation in the number of deaths over time. The fact that the NCI Atlas found counties with rates in the highest percentile in the rural North, while this present analysis did not, leads one to speculate about the role of chance in these findings. The Atlas reports the number of deaths in those counties falling in the fiftieth and sixtieth percentiles as 12 and 17 respectively.[125] It is quite obvious that even one or two excess deaths would alter the ranking of such counties. Using the NCI technique, counties with zero deaths are counted equally, regardless of population size. That is, zero deaths in a large, densely populated county is mapped exactly as is zero deaths in a large, sparsely populated county.

The OCER technique, by virtue of its division into equal risk categories, eliminates such problems. This technique enables one to test for systematic patterns of variation in mortality as opposed to the traditional method which simply highlights the counties with the highest rates. Given the instability of the mortality rates in a disease as rare as ovarian cancer, the OCER technique is obviously a useful tool.

---

[125]Hoover R., et.al. *op.cit.*, p.86

The findings of the geographic analyses using the OCER technique indicate a significantly higher mortality experience from ovarian cancer among the states in the central U.S., controlling for urbanization. It is possible that the difference in rates might be partially due to a higher case fatality rate among the central counties. Ideally, one would select a sample of the population in each area and follow them to determine the case fatality rates in the different areas.[126] Such a strategy is obviously not feasible in an ecological analysis such as this, but it is possible to control for a related factor in the analysis: access to medical care. Hypothesizing that those areas with better access to medical care would exhibit a lower case fatality rate due to earlier diagnosis and increased opportunity for therapeutic intervention, this variable was controlled for in subsequent geographic analyses using the summary variable described earlier. The lat-long index remained predictive of ovarian cancer mortality, and the results were strengthened after controlling for medical care access (chi square = 11.92; $p < .0001$). Access to medical care does act as a confounder in the association between ovarian cancer mortality and geographic variation as measured by the lat-long index, and its effect is to reduce the measured association. This result supports the hypothesis that the higher rates of ovarian cancer mortality in the central United States are due, in part, to poorer access to medical care which, in turn, suggests a higher case fatality rate.

[126]Lilienfeld A. *Foundations in Epidemiologic Research*. New York: Oxford University Press, 1976: 90.

6. SUMMARY AND CONCLUSIONS

6.1. Summary of results

Three primary findings emerged from this ecological investigation of ovarian cancer mortality in the United States. The first, described in Section 4.2, is the fact that different generations of women experience different risks of ovarian cancer. In other words, there is a cohort effect. The analysis conducted allowed for the separation of the influence of age, calander time, and birth cohort on ovarian cancer mortality from the years 1953 - 1978. The results indicate that the effect of birth cohort is intermediate to that of age and calander time and supports the idea that year of birth has an independent influence on a woman's lifetime risk of developing ovarian cancer. The analysis showed that those women of the birth cohorts of 1899-1903 and 1904-1908 experience greater risk of ovarian cancer mortality than women of earlier and later birth cohorts.

The second finding, described in Sections 5.1 and 5.2, is that birthrate in 1940 was found to be an important predictor of ovarian cancer mortality in 1970 in an ecological regression analysis of the predictive power of this and other variables. This effect was also found in an alternative method of analysis, the OCER technique, suggesting that the association is not artifactual, i.e.- an ecological fallacy.

The last finding of interest, described in Section 5.4, concerns the geographic patterns of mortality from ovarian cancer. The traditonal method of computer mapping of disease rates was compared to a technique in which the population was divided into equal risk categories on the basis of latitude (to test for east west variation), longitude (to test for north - south variation) and lat-long index (to test for variation from the center to the

borders of the United States) before analysis. The results of these two methods differed, pointing out the difficulties in evaluating the standard computer maps of county level morbidity or mortality rates and the likelihood of misinterpretation.

6.2. Epidemiologic limitations and suggestions for future research

There is still a great deal of work to be done to elucidate the epidemiology of ovarian carcinoma. The geographic patterns and time-trends of mortality rates have been well described, but the incidence of the disease, in general and by separate histologic type, has not. Accurate and timely diagnosis is a key factor in the understanding of any disease. The lack of same has greatly limited the research endeavors on cancer of the ovary. These tumors are difficult to ascertain in their early stages, resulting in a high case fatality rate and rendering estimates of incidence unreliable. Diagnostic methods are not applied uniformly within, much less among, medical care settings[127], exacerbating the problem of unreliability. Finally, most women with ovarian cancer are very ill by the the time of diagnosis, thus limiting their ability to participate in epidemiologic research endeavors. Progress in elucidating of the epidemiology of the disease relies, to some extent, on the improvement of diagnostic techniques, allowing for earlier detection and treatment. A recent international symposium in Venice revealed new diagnostic and therapeutic techniques which have begun to push the mean survival time beyond the 18 months reported earlier. These include: a new way of determining the spread of ovarian cancer which allows for more targeted treatment; and, refinements in the methods of delivering drugs directly to

---

[127]Barber op.cit., p.879.

the abdomen which allows for heavier doses of chemotherapy.[128]

Evaluating specific etiologic hypotheses of ovarian cancer is also problematic. The issue of ascertainment bias, raised earlier, is a potential confounder in any investigation of a risk factor that correlates with access to medical attention and the opportunity for diagnosis. Thus, sociodemographic variables such as income and degree of urbanization, observed to be predictive of ovarian cancer incidence and mortality, should be evaluated while controlling for access to medical care in order to determine their independent contribution to disease (See Sections 5.1 - 5.4).

In case-control studies of ovarian cancer patients and controls, there is a strong potential for recall bias in the information gathered, particularly in the area of reproductive and menstrual histories. It has been pointed out by several investigators[129][130][131][132][133] that patients with ovarian cancer may have a bias in recall of gynecologic events compared to control populations. Choice of controls should include women with benign ovarian diseases, women with other gynecologic disease (non-cancerous), and women from the general population. Comparisons between these control populations would help to elucidate any bias that might exist.

The role of pregnancy in reducing the risk of ovarian cancer has been well studied, but not with an eye to controlling the numerous potential confounders. In none of the nine case-control studies reviewed (See Section 2.2.1.2: Reproductive experience), was there a comprehensive effort to take account of the many variables that might affect the relationship between

[128]"*Ovarian Cancer*," New York Times (March 29, 1984), p.32.

[129]Barber, *op.cit.*

[130]Beral, *op.cit.*

[131]Lingeman, *op.cit.*

[132]Weiss, *op.cit.*

[133]Wynder, *op.cit.*

parity and ovarian cancer risk, i.e. - marital status, menstrual history, endocrinologic/hormonal profile, age at first pregnancy, reproductive history and contraceptive history. The findings of this ecological analysis add further support to the role of parity in protecting against ovarian cancer. The consistency of this finding in numerous different settings indicate that it is an important epidemiologic characterisitic, and future studies should attempt to uncover the underlying etiologic mechanism. These studies must take into account the potential confounders mentioned above, perhaps by matching cases and controls on contraceptive and reproductive histories.

Future research endeavors in the epidemiology of ovarian cancer must make use of a standardized classification scheme. Different histologic types of ovarian cancer might well result from different etiologies, and the practice of grouping all types of cancers together may mask important epidemiologic characteristics. Only if those conducting research on ovarian cancer utilize a common classification scheme can the potential for differing etiologies be assessed. This is paricularly important in comparing rates from other countries with those of the United States. Such comparisons can provide valuable epidemiologic leads, but only of one can be assured of relatively homogeneous histologic typing across the areas under consideration. Standardization of the classification scheme and the ascertainment efforts of cancer incidence registries such as the Third National Cancer Survey and the Surveillance, Epidemiology and End Results program should greatly improve our knowledge in this area.

Although researchers have gained some understanding of the role of pregnancy in reducing ovarian cancer risk, supported by the analyses presented in Sections 5.1 and 5.2 of this paper, other aspects of menstrual and reproductive life must be evaluated. The effect of environmental factors,

including diet, has been only superficially explored. As discussed in Sections 2.1 and 2.2.2, epidemiologic evidence suggests that environmental factors are of etiologic importance. The finding that the highest rates of ovarian cancer are in the most industrialized countries cannot be explained completely by ascertainment bias, particularly in light if the fact that highly industrialized Japan has one of the lowest rates in the world, and that Japanese migrants to the United States and their offspring show increased rates of the disease.[134] [135] [136] [137] Thus, it has been postulated that the causative factors must be more highly concentrated in the United States than in Japan.[138] A careful analysis of the geographic distribution and time trends of various hypothesized risk factors in the two countries would be an important contribution to the elucidation of their role in the etiology of ovarian cancer.

The stability of the mortality rates from ovarian cancer in the past two decades should not be taken as a rationale for ignoring the disease. This stability in the face of some improvement in diagnosis and treatment suggests that the disease is one that must be prevented because it is unlikely that it will ever be amenable to therapeutic intervention. As such, it is a prime candidate for further epidemiological research.

[134]Barber, *op.cit.*

[135]Lingeman, *op.cit.*

[136]Weiss, *op.cit.*

[137]Wynder et.al., *op.cit.*

[138]Barber *op.cit.*, p.36.

7. APPENDIX

7.1. Item A: Fortran program to calculate and create a data file of population estimates

```
        program pop4090
        implicit double precision (a-h,o-z)
c       interpolates population values between 1940-1990 using one of three
c       methods: piecewise linear, polynomial, or logarithmic polynomial
c       input file* disk$seedis001:[seedis.seedata.pop5080]
c       file 1 = pop4090.in1 or pop4090.t4 contains numeric data
c       file 2 = pop4090.in2 is original delist.dat
c       file 4 = pop4090.geo = pop4090.geo
c       file 3 (output) = pop4090.out
        character*2 xstate
        character*3 xcounty
        character*70 xtemp
        character*2 state(3500)
        character*3 herl(3500)
        character*2 sr(4)
        character*2 age(19)
        character*1 type(3)
        character*2 year(51)
        character*2 qtr(4)
        character*1 a
        character*2 b,c,d,e
        character*30 method(3)
        character*20 xsr(4)
        character*20 xage(19)
        character*10 quarter(4)
        character*2 ysr(3500)
        character*2 yage(3500)
        character*5 xyear(4)
        dimension isr(456)
        dimension iage(456)
        dimension itype(456)
        dimension iyr(456)
        dimension zyear(4),pop(4)
        dimension iqtr(456)
        data type/'1','2','3'/
        data sr/'WM', 'WF', 'NM', 'NF'/
        data age/'TO', '00', '05', '10', '15', '20', '25', '30', '35',
       *'40', '45', '50', '55', '60', '65', '70', '75', '80', '85'/
        data year/'40', '41', '42', '43', '44', '45', '46', '47', '48',
       *'49', '50', '51', '52', '53', '54', '55', '56', '57', '58','59',
       *'60', '61', '62', '63', '64', '65', '66', '67', '68','69',
       *'70', '71', '72', '73', '74', '75', '76', '77', '78','79',
       *'80', '81', '82', '83', '84', '85', '86', '87', '88','89', '90'/
        data qtr/ '00', '25', '50', '75'/
        data method/'piecewise linear', 'polynomial',
       *'logarithmic polynomial'/
```

```
       data xsr/'white males', 'white females', 'non-white males',
      *'non-white females'/
       data xage/'all ages','0-4 years','5-9 years','10-14 years',
      *'15-19 years','20-24 years','25-29 years','30-34 years',
      *'35-39 years','40-44 years','45-49 years','50-54 years',
      *'55-59 years','60-64 years','65-69 years','70-74 years',
      *'75-79 years','80-84 years','85 years and over'/
       data xyear/'50.25','60.25','70.25','80.25'/
       data zyear/50.25,60.25,70.25,80.25/
       data quarter/'1 January', '1 April', '1 July', '1 October'/
c      read file 2 completely (pop4090.in2)
       open(unit=2,type='OLD',readonly)
       nde=0
  2    continue
       nde=nde+1
       read(2,200,end=299)a,b,c,d,e
200    format(3x,a1,1x,a2,1x,a2,1x,a2,1x,a2)
       do 21 j=1,3
       if (a.eq.type(j)) itype(nde)=j
  21   continue
       do 22 j=1,4
       if (b.eq.sr(j)) isr(nde)=j
  22   continue
       do 23 j=1,19
       if (c.eq.age(j)) iage(nde)=j
  23   continue
       do 24 j=1,51
       if (d.eq.year(j)) iyr(nde)=j
  24   continue
       do 25 j=1,4
       if (e.eq.qtr(j)) iqtr(nde)=j
  25   continue
       go to 2
299    nde=nde-1
       close(unit=2)
c      read file 4 (state level geocode file) completely
       open(unit=4,type='OLD',readonly)
  4    continue
       read(4,400)xtemp
400    format (a70)
       if (xtemp(1:7).ne.'END DDF') go to 4
c end  of ddf on geocode file
       nareas=0
c store geocodes for entire file
  41   continue
       nareas=nareas+1
       read(4,401,end=499) state(nareas)
401    format(a2)
       go to 41
499    continue
c end of geocode file reached
       nareas=nareas-1
       close(unit=4)
```

```
c     write ddf for output file
      open(unit=3,type='NEW',carriagecontrol='LIST')
      nde2=nde+1
      write(3,300)nde2,nareas
300   format('nde=',i8/'areas=',i8/'card=70'/'*LEVEL = STATE')
      write(3,301)
301   format('DE=FIPS.STATE'/'  T=A'/'  U=K'/'  L=2'/'   S=1'/
     *'  h=#FIPS STATE CODE#')
      do 10 j=1,nde
      i1=itype(j)
      i2=isr(j)
      i3=iage(j)
      i4=iyr(j)
      i5=iqtr(j)
      write(3,310)type(i1),sr(i2),age(i3),year(i4),qtr(i5)
      istart=(j*70)+1
310   format('DE=POP',a1,'.',a2,'.',a2,'.',a2,'.',a2)
c     write headers
      write(3,311)istart
311   format('  T=I'/'  U=D'/'  L=12'/'S=',I10)
      write(3,312)
312   format('  h=#Estimated population#')
      write(3,313)method(i1)
313   format('  h=#',a30,'#')
      write(3,314)xsr(i2)
314   format('  h=#',a20,'#')
      write(3,315)xage(i3)
315   format('  h=#',a20,'#')
      write(3,316)quarter(i5),year(i4)
316   format('  h=#',a10,'19',a2,'#')
 10   continue
      write(3,317)
317   format('end ddf')
c     read file 1 completely and write df of file 3
      open(unit=1,type='OLD',readonly)
c loop over areas
      do 399 i=1,nareas
      write(3,351) state(i)
351   format(a2)
c loop over data elements
      do 398 j=1,nde
      imeth=itype(j)
      read(1,100) pop
100   format(4f8.0)
      ktimes=4
      do 3515 jj=1,4
c pop -999 means missing, reset ktimes
      if (pop(jj).eq.-999) ktimes=ktimes-1
c if logarithmic polynomial set any non-positive pop to 0.1
      if(pop(jj).eq.0.and.imeth.eq.3) pop(jj)=0.1
3515  continue
c     calculate population estimates using method 1,2 or 3
      yearout=(38.75+iyr(j)+iqtr(j)*.25)
```

```
       call oneyear(imeth,ktimes,zyear,pop,yearout,popout,ier)
       if(popout.lt.0)popout=-999
       write(3,352)popout
352    format(f12.0)
398    continue
399    continue
       close(unit=3)
       call exit
       end
       subroutine oneyear(imeth,ktimes,zyear,pop,yearout,popout,ier)
       implicit double precision (a-h,o-z)
       dimension zyear(1),pop(1)
       dimension temp(4),c(4)
c if logarithmic polynomial set any non-positive pop to 0.1
       if(pop(jj).eq.0.and.imeth.eq.3) pop(jj)=0.1
       temp(1)=pop(1)
       temp(2)=pop(2)
       temp(3)=pop(3)
       temp(4)=pop(4)
       go to (10,21,20),imeth
10     call linex(zyear,pop,yearout,popout)
       return
c if logarithmic polynomial set any non-positive pop to 0.1
20     if(temp(1).eq.0.and.imeth.eq.3) temp(1)=0.1
       temp(1)=log(temp(1))
       if(temp(2).eq.0.and.imeth.eq.3) temp(2)=0.1
       temp(2)=log(temp(2))
       if(temp(3).eq.0.and.imeth.eq.3) temp(3)=0.1
       temp(3)=log(temp(3))
       if(temp(4).eq.0.and.imeth.eq.3) temp(4)=0.1
       temp(4)=log(temp(4))
21     call coeff(zyear,temp,ktimes,ier,c)
       poptmp=0
       do 6 kk=1,ktimes
       poptmp=poptmp+c(kk)*yearout**(kk-1)
6      continue
       popout=poptmp
       if(popout.le.0.and.imeth.eq.3) popout=-999
       if(imeth.eq.3) popout=exp(popout)
       return
       end
       subroutine linex(zyear,pop,yearout,popout)
       implicit double precision (a-h,o-z)
       dimension pop(1),zyear(1)
       if(pop(1).le.1) return
       if(yearout.ge.zyear(2)) go to 10
       popa=pop(2)
       popb=pop(1)
       spread=zyear(2)-zyear(1)
       temp=(zyear(2)+zyear(1))/2
       go to 99
10     continue
       if(yearout.ge.zyear(3)) go to 11
```

```
          popa=pop(3)
          popb=pop(2)
          spread=zyear(3)-zyear(2)
          temp=(zyear(3)+zyear(2))/2
          go to 99
   11     popa=pop(4)
          popb=pop(3)
          spread=zyear(4)-zyear(3)
          temp=(zyear(4)+zyear(3))/2
   99     continue
          popout=(popa+popb)/2+((popa-popb)/spread)*(yearout-temp)
          return
          end
          subroutine coeff(x,y,k,ier,b)
          implicit double precision (a-h,o-z)
          dimension x(1),y(1),b(1),s(4,4)
cc  ier=1 normal condition  ier=-1 error condition
          do 10 i=1,k
          do 10 j=1,k
          s(i,j)=x(i)**(j-1)
   10     continue
          call mat(s,k,ier)
          do 11 i=1,k
          b(i)=0.0
          do 11 j=1,k
          b(i)=b(i)+s(i,j)*y(j)
   11     continue
          return
          end
          subroutine mat(a,n,ier)
          implicit double precision (a-h,o-z)
          dimension a(n,n)
          ier=1
          error=0.000001
          do 1 i=1,n
          if(abs(a(i,i)).gt.error) go to 10
          ier=-1
          return
   10     continue
          a(i,i)=1.0/a(i,i)
          do 2 j=1,n
          if(j.eq.i) go to 2
          a(i,j)=a(i,j)*a(i,i)
   2      continue
          do 1 j=1,n
          if(j.eq.i) go to 1
          do 6 k=1,n
          if(k.eq.i) go to 6
          a(j,k)=a(j,k)-a(j,i)*a(i,k)
   6      continue
          a(j,i)=-a(j,i)*a(i,i)
   1      continue
          return
          end
```

7.2. Item B: United States fertility rates, 1925 - 1960

ITEM B: FERTILITY RATES, UNITED STATES: 1925-1960; (SELECTED YEARS); WHITES AND NON-WHITES."		
YEAR	FERTILITY RATE (Births per 1,000 women aged 15-44)	
	White	Non-white
1925	103.3	134.0
x	x	x
x	x	x
x	x	x
x	x	x
1930	87.1	105.9
x	x	x
x	x	x
x	x	x
x	x	x
1935	74.5	98.4
1936	73.3	95.9
1937	74.4	99.4
1938	76.5	100.5
1939	74.8	100.1
1940	77.1	102.4
1941	80.7	105.4
1942	89.5	107.6
1943	92.3	111.0
1944	86.3	108.5
1945	83.4	106.0
1946	100.4	113.9
x	x	x
x	x	x
x	x	x
1950	102.3	137.3
x	x	x
x	x	x
x	x	x
x	x	x
x	x	x
x	x	x
x	x	x
x	x	x
x	x	x
1960	113.2	153.6

7.3. Item C: Ovarian cancer mortality, United States, 1968 - 1978

TRUNCATED CRUDE RATES; WOMEN AGE >35;
WHITES AND NON-WHITES; 1968 - 1978.

	1968	1969	1970	1971	1972	1973	1974	1975	1976	1977	1978
AL	7.1	6.7	6.3	6.1	7.7	7.1	8.2	8.5	8.4	7.5	8.4
	6.7	5.5	5.0	5.7	6.9	5.7	4.0	5.7	6.5	4.8	6.5
AK	1.5	0.7	1.5	3.9	6.3	4.7	2.3	0.7	6.3	3.9	3.9
	0.0	0.0	0.0	2.7	0.0	2.7	0.0	0.0	2.7	5.5	8.3
AZ	5.7	5.1	6.4	7.1	6.8	7.7	7.2	8.7	8.7	7.9	11.1
	0.7	2.3	2.3	1.5	3.1	2.3	3.9	4.7	2.3	1.5	2.3
AR	6.4	9.5	7.9	7.2	5.9	7.5	9.7	9.5	8.2	8.7	10.6
	6.9	3.9	5.4	7.4	8.9	7.4	4.4	3.9	4.4	4.4	5.9
CA	8.0	8.6	9.6	9.3	9.3	8.9	9.5	9.8	10.5	9.7	10.8
	3.1	3.0	3.1	2.4	3.0	2.3	3.3	3.7	4.6	4.7	4.5
CO	7.0	6.7	7.6	8.4	6.2	7.0	6.2	8.5	8.2	7.9	8.7
	1.2	1.2	1.2	0.0	2.5	2.5	1.2	1.2	2.5	0.0	1.2
CT	9.5	9.8	10.2	10.7	7.9	10.2	9.9	11.1	11.6	9.7	8.7
	0.7	3.9	4.7	2.3	1.5	2.3	3.9	3.1	6.2	3.1	3.1
DE	8.1	5.7	10.4	11.1	8.4	8.1	10.0	7.7	8.4	10.4	10.0
	0.0	2.0	4.0	6.1	4.0	0.0	6.1	4.0	0.0	8.1	2.0
DC	20.9	20.0	24.5	18.2	23.6	19.1	20.0	15.4	17.3	10.0	11.8
	6.6	6.9	8.2	4.3	5.9	9.6	9.2	3.9	4.9	2.9	6.6
FL	9.0	9.6	9.7	10.7	10.1	9.7	12.3	11.9	13.8	14.6	15.1
	3.6	4.7	4.2	4.7	3.9	6.2	6.0	5.4	6.3	5.4	3.9
GA	6.0	5.8	6.1	6.8	8.0	6.0	7.8	7.6	7.7	6.4	8.4
	3.7	4.6	4.7	4.3	5.0	5.0	4.0	4.7	4.3	4.9	5.4
HI	2.6	5.2	1.9	4.5	2.6	6.5	5.2	1.9	0.6	7.2	5.2
	5.5	2.9	4.4	3.3	5.1	4.4	5.5	2.5	5.1	3.6	6.2
ID	7.0	6.0	6.0	6.0	8.0	5.0	7.2	8.2	8.2	9.0	9.2
	0.0	0.0	10.0	0.0	0.0	0.0	0.0	10.0	20.0	0.0	0.0
IL	9.5	11.2	10.2	9.4	11.2	10.4	11.0	10.6	10.6	10.6	10.5
	4.5	4.3	5.5	4.9	3.3	4.3	3.6	4.8	4.4	4.9	5.3
IN	8.4	8.3	8.4	9.1	7.6	7.6	8.5	8.7	9.6	8.2	8.9
	2.6	2.6	5.3	3.5	2.6	3.5	4.9	1.7	7.6	5.8	3.5

```
IA   11.2 10.1 10.7 10.1 10.5 11.4 10.5 10.5 10.5  9.4 11.2
      3.7  0.0  0.0  7.5  0.0  7.5  7.5  0.0  3.7  0.0  3.7

KS    8.5  9.1  8.3 10.2 10.9  8.9  9.5  8.2 10.3 10.4  9.4
      5.3  6.6  2.6  4.0  0.0  8.0  5.3  2.6  6.6  2.6  2.6

KY    6.9  7.4  8.1  7.8  9.7  7.3  9.0  8.2  9.3  7.5  9.0
     13.3  8.1  8.8  6.6 13.3  5.9  7.4  6.6 10.3  8.8  2.9

LA    5.6  4.5  5.0  4.7  5.7  5.2  5.7  6.2  6.5  6.5  6.0
      3.7  4.7  3.7  4.1  3.1  3.9  2.8  2.9  4.7  2.8  3.9

ME   11.2  9.6  9.9  9.2 12.6  8.1 10.1  9.6 10.5 12.3  9.2
      0.0  0.0  0.0 23.7  0.0  0.0  0.0 23.7  0.0  0.0  0.0

MD    8.0  7.8  6.9  6.7  8.6  7.4  9.6  8.0  8.9  9.2  8.6
      3.3  4.8  3.9  4.4  5.2  4.1  5.7  5.2  4.1  6.4  4.6

MA    9.1 10.9 10.7 11.1 11.3  9.8 10.7 10.9 10.2 10.1 10.0
      2.1  6.3  1.4  6.3  0.0  3.5  4.9  4.9  2.1  4.9  2.1

MI    8.3  7.5  8.5  8.7  8.5  7.8  8.6  9.1  8.9  9.1  8.8
      5.9  3.5  4.3  5.2  5.7  5.1  4.7  5.5  5.9  3.8  3.9

MN    8.9  9.9  8.6  9.6 10.6  9.7  8.7  9.5  8.8 10.9  8.3
      2.1  0.0  0.0  4.3  0.0  0.0  2.1  0.0  2.1  6.4  2.1

MS    5.0  6.2  5.8  6.7  7.3  5.5  8.6  7.4  6.4  7.3  6.4
      4.5  5.6  5.0  4.7  4.7  4.1  5.6  4.5  4.7  4.7  6.9

MO    9.4  9.8  9.6  9.5 11.4  9.8 10.5 10.6  8.6 10.1 10.6
      3.4  7.5  2.7  8.5 10.3  6.5  5.1  5.1  7.2  6.5  6.1

MT   10.0  5.0  6.9  4.7  4.4  6.3  7.7  8.0  8.3  9.4  9.1
     10.8  0.0  0.0  0.0  0.0  5.4  0.0  0.0  0.0  0.0  5.4

NE    9.7 12.3 11.4  9.4 13.2 10.4  7.6  9.9 12.7 10.6 12.0
     15.8  6.3  6.3  0.0  0.0  6.3  3.1  3.1  3.1  3.1  9.5

NV    3.6  6.5  5.4  8.6  6.5  7.2  6.5  7.2 10.5  9.0  7.9
      0.0  3.3  0.0  3.3  0.0  0.0  3.3  3.3 10.1  0.0  0.0

NH    7.4 10.0  9.5 10.0  8.8 10.7 10.0 10.5 12.6 11.0 12.4
      0.0  0.0  0.0  0.0  0.0 33.2  0.0  0.0  0.0  0.0  0.0

NJ   10.9 10.9 10.4 11.0 10.2 11.5 11.3 11.6 11.9 12.1 12.2
      3.4  3.0  4.0  4.5  8.7  7.8  4.5  4.3  3.6  4.5  4.2

NM    6.0  4.8  5.8  6.8  6.8  7.2  8.2  7.0  7.2  5.4  8.2
      0.0  1.2  0.0  2.4  2.4  1.2  1.2  0.0  0.0  4.8  2.4

NY   11.2 11.6 12.2 11.0 11.3 11.0 11.4 11.2 12.0 11.1 10.8
      4.1  3.6  4.5  4.8  3.7  3.8  4.7  4.1  4.5  4.2  4.7
```

NC 5.9 5.4 6.6 7.2 6.4 8.0 7.1 7.9 7.1 8.2 8.5
    3.5 3.2 4.1 2.8 5.6 5.4 3.5 4.5 5.3 5.1 5.6

ND 6.6 9.2 9.5 10.4 8.2 10.4 15.8 10.4 7.9 9.5 11.1
    9.3 0.0 0.0 0.0 0.0 18.6 0.0 0.0 9.3 0.0 0.0

OH 8.6 9.6 9.5 9.5 9.2 10.0 9.6 9.6 9.1 9.4 9.5
    5.4 3.0 5.2 3.7 6.1 4.9 6.4 4.5 5.4 4.2 6.9

OK 7.3 7.4 9.3 8.4 6.8 7.9 9.0 8.9 10.4 8.7 7.8
    4.0 5.7 5.7 6.3 5.7 4.5 7.4 6.8 5.7 5.1 4.0

OR 9.3 8.4 9.4 9.4 8.0 9.1 8.6 9.7 8.9 11.7 11.5
    2.3 2.3 0.0 2.3 19.1 0.0 7.1 4.7 9.5 0.0 2.3

PA 9.8 10.7 9.4 9.9 10.2 10.2 9.1 10.6 10.9 10.3 10.9
    5.8 5.6 5.8 6.6 6.4 7.6 5.8 5.5 5.3 6.2 6.8

RI 8.6 7.0 10.2 10.2 12.7 10.2 9.4 11.6 11.8 12.4 10.2
    5.1 10.3 10.3 0.0 0.0 0.0 0.0 5.1 10.3 5.5 5.1

SC 5.2 6.8 7.4 6.1 6.3 7.2 7.5 6.6 7.4 6.6 7.0
    3.4 3.4 2.4 4.3 3.4 6.5 4.3 2.4 4.3 4.3 3.4

SD 10.7 9.2 11.0 10.4 13.1 11.3 8.0 10.7 12.2 9.5 10.7
    0.0 9.5 14.2 9.5 9.5 0.0 9.5 0.0 4.7 0.0 0.0

TN 8.4 6.8 7.7 7.1 7.0 8.4 8.4 8.7 8.1 9.3 8.4
    7.0 5.6 4.5 4.5 4.8 6.2 9.7 5.9 7.0 6.2 4.5

TX 6.9 6.9 6.7 7.3 6.7 7.2 7.2 7.3 7.7 8.4 7.2
    4.0 5.1 4.8 4.7 4.1 4.1 3.8 4.5 4.6 4.8 4.4

UT 5.1 5.4 5.4 4.4 5.6 5.7 4.6 5.1 4.4 8.7 7.4
    0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 4.9

VT 9.5 7.1 8.7 10.3 7.9 10.3 7.1 9.1 12.7 10.3 7.9
    0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0

VA 6.6 7.4 7.0 6.9 8.0 7.9 7.8 8.5 8.4 6.6 7.6
    4.6 5.6 4.3 5.2 4.6 3.3 6.0 4.8 5.6 5.8 7.2

WA 8.1 9.0 8.9 8.6 8.9 10.5 8.5 9.2 9.6 9.8 9.3
    3.7 1.8 4.6 2.8 5.6 1.8 6.5 2.8 1.8 3.7 1.8

WV 8.8 7.2 7.4 9.4 9.4 7.9 8.7 8.5 6.9 8.2 8.9
    0.0 2.5 12.9 5.1 10.3 10.3 2.5 15.5 12.9 7.7 18.0

WI 10.0 10.0 11.0 9.3 9.9 9.6 10.6 10.6 10.6 10.4 10.2
    0.9 0.9 0.9 2.9 1.9 1.9 6.8 3.9 0.0 2.9 4.8

WY 7.6 7.6 5.4 6.5 6.5 6.5 5.4 2.7 11.4 7.6 5.9
    0.0 0.0 0.0 0.0 0.0 15.6 0.0 0.0 0.0 0.0 0.0