

# RELATION OF CARCINOMA IN SITU TO INVASIVE CARCINOMA OF THE CERVIX UTERI

JOHN E. DUNN, JR.  
NATIONAL CANCER INSTITUTE

## 1. Introduction

The quantitative measurements used to express the rate of occurrence of disease and the magnitude of its accumulation are, respectively, incidence and prevalence. Incidence is the number of cases of a disease being produced by a population over time and is expressed as a rate of production per unit of time per unit of population. Prevalence, on the other hand, is the number of persons in the population having the disease at any moment in time, and as a rate is expressed as the number of cases per unit of population. Prevalence of disease, then, may be looked upon as accumulated incidence.

## 2. Interrelationship of incidence and prevalence of disease

Incidence of disease is the prime interest of the epidemiologist since he is concerned with determining the probability in some period of time of a nondiseased individual within a defined population developing the disease in question. Since the state of the same individual without disease and with disease is a function of time, incidence of disease, like prevalence, also has its relationship to a moment in time. However, with incidence this moment is the point of time in the evolution of the disease when the disease can be said to exist. The implication is clear from this that the distinction can be made fairly sharply in disease development time when an individual is free of and when he is a victim of a disease. For diseases of identifiable, exogenous origin that run an acute course this is not a particularly important problem. But for diseases of unknown or uncertain causation that run an insidious and protracted course the point in the disease development time when disease is said to have its beginning may be very arbitrary and artificial.

It has been said that there are no diseased people but only sick people. Medicine has developed a system for categorizing sick people into diseased people. To paraphrase Humpty Dumpty, our labels for disease mean exactly what we choose they shall mean—nothing more and nothing less. Our concept of disease has traditionally been that we start out with sick people and the problem of

clinical medicine has been to sort them into defined categories with disease labels. As medicine has progressed from its largely empirical basis to a physiological, biochemical, and pathological understanding of disease, the point in disease development time when a disease can be said to be present is changing. And with the development of the means of measuring abnormal deviations of biochemical and physiologic function and relating these to disease processes it becomes possible to identify disease among persons not considered sick as well as among the sick.

If we consider disease schematically we might represent it as a disease develop-

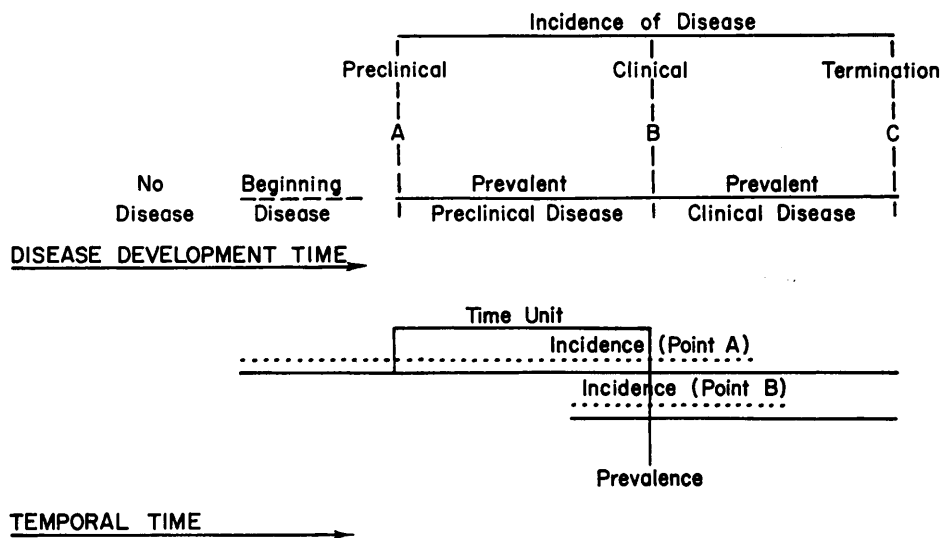


FIGURE 1

Schematic representation of incidence and prevalence of disease in relation to disease development time and temporal time.

ment time line, as in the upper half of figure 1. Here is shown the progress of disease from its indefinable beginning in the well person to point *A* where, if we have the means, we may be able to say the disease is present although the person feels well. Next, progress continues to point *B* where the person is sick and the clinical syndrome characterizing the disease is present. And finally, to point *C* where the patient recovers from or dies from his disease.

Incidence of this disease can be determined equally well whether we identify it as it appears in a population at points *A*, *B*, or *C*. It is essential that only one point be used, however, and that the pattern of events that bring patients to a diagnosis be fairly well stabilized. For example, if some in a population were having their disease diagnosed at *A* as well as all persons that were at *B*, the apparent incidence rate would be in excess of the true incidence rate. Similarly,

if the pattern for diagnosis at *B* were disturbed, a biasing effect on the incidence rate would occur. For example, if women were exhorted to practice breast self-examination, and breast nodules were discovered earlier as a result than they would have been discovered by accident, as had been the pattern before, breast cancer incidence would appear to be increased.

The ability to diagnose disease at point *A* in development time, as well as at *B*, is much more disturbing to the determination of true incidence than the statement made above would indicate. If the diagnostic procedure can diagnose the disease at point *A*, then it will undoubtedly be able to diagnose the disease at any point in its development time between *A* and *B*. Among those examined by the procedure that can make a diagnosis at *A*, the prevalent preclinical disease will be discovered. Determination of true incidence will no longer be possible unless all disease as it develops is now diagnosed at *A*, or until the deficit at *B* is overcome through abandoning diagnosis between *A* and *B* and allowing enough time to elapse for all cases found between *A* and *B* to have reached point *B*.

In the lower half of figure 1, use of the procedure that can make the diagnosis at point *A* is represented in relation to temporal time. When applied to a population for the first time it would discover all disease in the population that had reached point *A* in its development and beyond, less the disease already identified by conventional diagnosis that had reached point *B* and beyond. This would be the prevalence of all developing disease lying between *A* and *B*.

To summarize then, incidence of disease, as illustrated in the upper half of figure 1, is a point in disease development time, and prevalence is an interval of this time. In temporal time, as illustrated in the lower half of figure 1, prevalence of disease is at a point in temporal time and incidence represents an interval of such time.

### 3. Disease-oriented rather than population-oriented epidemiology

It is the usual practice to consider the chronic diseases in terms of an actual population. Where one is considering the competition between diseases within a single population this is essential. For comparisons of a single disease between total populations this preoccupation with expressing summary rates in terms of a population leads to some artificiality. Most chronic diseases increase markedly with age. To compare the occurrence of a disease between two total populations, corrections must be made for any differences in the age distribution of the populations. This is done by weighting the age-specific rates for the populations to be compared with a common set of age-specific population weights. These may be derived from one or the other of the populations being compared or from an unrelated population.

In table I, age-specific incidence rates for cervical cancer and the percentage distribution of the population that gave rise to them is shown. The summary rate for cervical cancer in this population gives almost three times the weight

to the age group where the rate is the lowest, as it does to the oldest age group where the rate is the highest. Comparison of this population with another population might involve weighting these age-specific rates with a different set of population proportions and the summary rate would no longer express the experience of an actual population. A further complication results from the fact that there is no unanimity as to the "standard" population for age adjustment so that interstudy comparisons are often not possible.

TABLE I

POPULATION PERCENTAGE DISTRIBUTION AND INCIDENCE OF CERVICAL CANCER PER 100,000 POPULATION FOR WHITE WOMEN, WESTERN REGION, 1947 [6]

	Age Groups					
	20-29	30-39	40-49	50-59	60-69	70+
Population	22.6	22.1	19.4	16.0	11.8	8.2
Rates	5.9	33.1	72.2	98.9	86.1	95.2

One can take the position that the age composition of a population is only a reflection of the vicissitudes of life and that the morbid force of a single disease might be considered apart from this. Such would be the case if we expressed the summary rate as an unweighted average of the age-specific rates. Such a rate, then, might be defined as the maximum incidence of the disease as it would appear in a rectangular population with equal numbers in each age group of the population. Interpopulation comparisons would be simplified with such summary rates. Orientation now would be primarily with the disease rather than with populations and their age composition.

There are advantages in orienting oneself with the disease and regarding a population as the medium through which the disease expresses itself. For this purpose we might wish to study an age cohort of population from birth to death and observe its lifetime experience with a disease. The only deficiency here, and one that is unavoidable, is that the cohort is continually dying off as it ages and our observations would be limited to those surviving the continuous attrition of general mortality. While such a plan of study is perhaps impractical, age cohorts can be observed historically through comparable periods of their life span with revealing results. The study of lung cancer and tuberculosis among men are examples of the usefulness of this method.

Another orientation one can make with the disease is that represented in the upper half of figure 1. The complete genesis of the chronic noninfectious diseases that are constantly occurring in a population is available at all times if we but have the means to reveal it. Referring to figure 1, assume we have the means to diagnose a disease at *A*, whereas diagnosis previously had been at *B*. If this new method were applied to a population, then all disease

would be identified that is at any point in its genesis between *A* and *B*. If we now continued to re-examine at periodic intervals the nondiseased population we would reveal all new disease reaching point *A* in the re-examination intervals. This would be the incidence of new disease being produced by the population and might be termed a disease cohort. The prevalent disease found on first examination, then, represents a succession of disease cohorts moving from point *A* to point *B* in disease genesis. Since we know how many of these there are, and we have determined the size of each new cohort added in an interval of time, the ratio of prevalence to incidence will give an estimate of the time interval separating point *A* and *B*. As successive disease cohorts move in time from *A* to *B*, there will be continuous attrition from general mortality, as is true for all aging populations. We can avoid this bias by expressing incidence and prevalence as age-specific rates with any summary rate derived from the unweighted age-specific rates. As stated previously, this is equivalent to studying the genesis of a disease as though it were occurring in a rectangular population in which there are equal numbers in each age group of the population. Further, by observing the age-specific incidence and prevalence rates over successive age groups of the population some conclusions can be reached as to whether the disease is behaving differently when it has its origin at different ages. Such differences, if observed, may be due to a characteristic of the disease itself or a reflection of a change in disease experience for age cohorts.

#### 4. Disease-oriented epidemiology of cancer of the cervix uteri

The specific disease I shall consider in this way is cancer of the cervix uteri. In 1928 Papanicolaou [1] first reported that he could identify cancer cells in vaginal fluid taken from a woman having the disease. This observation went largely unnoticed for a number of years until collaboration with Traut, a gynecologist, demonstrated its clinical usefulness and this was brought to the attention of the medical profession in a monograph published in 1943 [2]. As cytology became more widely used as an examination procedure it became clear that cervical cancer could be diagnosed before the patient was aware of any symptoms and before the disease was suspected by clinical examination. Further than this, many of the lesions when examined histologically failed to show invasion of the supporting tissue by the cancerous appearing epithelium, the classical criterion of malignancy. The question raised was, are these really malignant lesions?

An obvious means of answering this question would appear to be to identify lesions of the kind in question, which have been given many names including that most commonly used—carcinoma in situ—which would seem to deny the question, and follow them to see what their eventual fate might be. I shall not go into the several problems in this plan of study, which have been discussed elsewhere [3], except to say that one cannot at the same time remove a tissue to see what it is and leave it undisturbed to see what it will become.

An epidemiological approach to answer the question of the nature of carcinoma in situ was proposed [3] based on the concepts depicted in figure 1. The disease development time that would be revealed from examining a population of women cytologically would be that shown in figure 2.

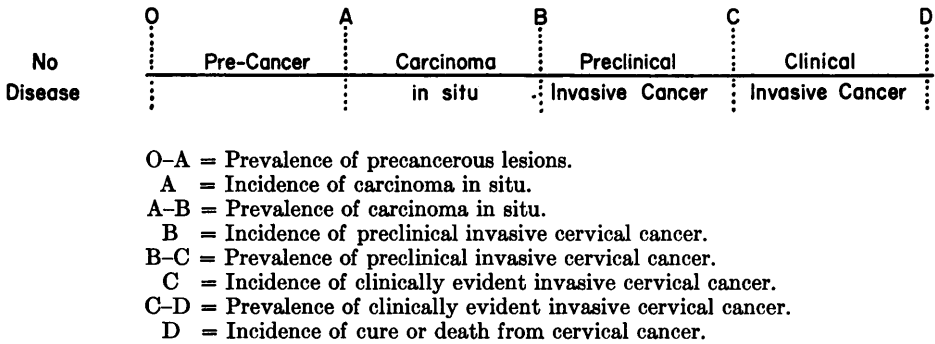


FIGURE 2

Disease development time for cervical cancer.

Under the usual conditions of diagnosing disease, the cervical cancer known to exist in a population is represented by the interval from point *C* to point *D*. Examining a population of women with cytology, and tissue diagnosis of those with suspicious cytology findings, would in addition identify nearly all with cervical cancer anywhere in the interval from point *A* to *C*. Those in the interval between *A* and *B* can be distinguished histologically from those in the interval between *B* and *C*. Some portion of those between points *O* and *A* would also be identified although not clearly recognized. Attention will be limited here to the disease development interval represented by *A* to *C*.

In column 5 of Table II are shown the age-specific incidence rates for clinically recognized invasive cervical cancer occurring in Memphis and Shelby County, Tennessee, before a community-wide use of cytology was made. In columns 2 and 4 are shown the prevalence rates for carcinoma in situ and unsuspected invasive cervical cancer found in the subsequent cytological screening of 53,585 white women [5].

In terms of figure 2, the age-specific rates of column 2 of table II represent the interval *A* to *B*; those of column 4, the interval *B* to *C*; and those of column 5, the point marked *C*. It is not practicable to obtain observational data for point *B*; that for point *A* will come from periodic re-examinations of women found cytologically negative on first examination.

We can view the prevalence rates of columns 2 and 4 of table II as pools of disease that change in magnitude over the ages of women. These pools are continually being fed with new disease that is developing, and being depleted by maturing disease that is progressing to the next stage. The prevalence rates of

TABLE II

OBSERVED AND CALCULATED PREVALENCE AND INCIDENCE RATES PER 1,000 WHITE WOMEN FOR CARCINOMA IN SITU AND INVASIVE CANCERS OF THE CERVIX—MEMPHIS AND SHELBY COUNTY, TENNESSEE [4]

(Rates in parentheses are based on less than five cases)

Age Group	Calculated Incidence Carcinoma in Situ (1)	Observed Prevalence Carcinoma in Situ (2)	Calculated Incidence Preclinical Invasive Cancer (3)	Observed Prevalence Preclinical Invasive Cancer (4)	Incidence Clinical Invasive Cancer (5)
20—	0.5	2.4	0.02	(0.1)	0.0
30—	0.7	4.8	0.4	1.3	0.3
40—	0.8	4.2	0.8	1.6	0.8
50—	1.0	2.8	1.2	3.2	1.0
60—	1.7	5.6	1.4	5.6	1.2
70—	1.4	(3.6)	1.6	(7.3)	1.4

column 4, for example, are being depleted in the corresponding age groups of women at the rates shown in column 5, and being replenished at the unobserved rates represented by column 3. Since the observed rates of column 4 are the net result of outflow as observed in column 5, and inflow at the unknown rates of column 3, we are in a position to arrive at some estimates of the latter. The rates of columns 4 and 5 can be taken to be estimates for the middle year of the age group interval. For age group 20–29 there is negligible outflow from the prevalence pool of preclinical invasive cervical cancer as clinical disease (column 5). The incidence of preclinical invasive cervical cancer before age 20 is probably negligible. The prevalence rate of 0.1 per 1,000 women at age 25 would then result from an average annual inflow of 0.02 per 1,000 for each single year of age for women progressing from 20 to 25 years of age. By age 35 the prevalence pool of column 4 had increased to 1.3 per 1,000. This represents a net annual increment between inflow and outflow of 0.12 per 1,000 women  $[(1.3 - 0.1)/10]$  for women going from age 25–35. At age 35 the outflow (column 5) is 0.3 per 1,000. To satisfy this outflow, and the increment of increase of the prevalence pool, would require an incidence rate of 0.4 per 1,000  $(0.3 + 0.12)$  at this age for preclinical invasive cancer. The remaining rates of column 3 were calculated in a similar fashion.

The calculated rates of column 3 are of academic interest only since determining these rates from observational data is not feasible. The calculated incidence rates of carcinoma in situ of column 1, however, can be determined from the rates of columns 2 and 3, as were the rates of column 3 from those of columns 4 and 5. A more direct procedure would be to combine the age group prevalence rates of columns 2 and 4 as representing the prevalence pool of disease lying between points *A* and *C* of figure 2. Outflow from this pool is at the rates of column 5 and the inflow rates of column 1 are determined as before. The calculated rates of carcinoma in situ incidence of column 1 are based on the

assumption that carcinoma in situ is the beginning of all cervical cancer, and all in situ carcinoma eventually progresses to the invasive disease.

As stated earlier, the ratio of prevalence to incidence for a stage of disease is a measure of the average time spent by cases of disease between entrance to and exit from, the prevalence pool. From table 2 the ratio of the summed rates of columns 4 and 5 ( $19.1/5.4$ ) is 3.5. The average duration of preclinical invasive cervical cancer for a population with equal numbers in each age group, then, would be approximately 3.5 years. There would be some variation around this average. At age 35, for example, the incidence rate for clinical cancer is 0.3 per 1,000 and increasing approximately 0.05 for each additional year of age. These rates would exhaust the 1.3 per 1,000 prevalence cases present at age 35 in about 3.5 years. At this time the cases that entered the prevalence pool at age 35 would have been in the prevalence pool the longest time and on the verge of withdrawal to the next stage. At age 45 the duration computed in this way would average about two years and at age 65 between four and five years.

Similarly, for carcinoma in situ the average duration would be the ratio of the sums of rates of column 2 to column 1 ( $23.4/6.1$ ) for an estimate of average duration of 3.8 years. The variation around this average over the ages of women is even greater than for preclinical invasive cancer. Considering again the time necessary to exhaust the prevalence pool at different ages, the duration would be 8–10 years for women developing carcinoma in situ in the 20's and 30's; shortening to 2–3 years in the 50's; and increasing somewhat again at older ages. These computations, of course, are based on the assumption that carcinoma in situ invariably progresses to invasive cancer.

Of the women examined cytologically for the first time with the results as given in table 2, a second and third cytological screening examination was done on 15,929. The findings from these examinations in terms of person-years of observation and the cervical cancers found are given in table III. There were 20 carcinomas in situ produced by this population during the time it was under observation. If we apply the age-specific incidence rates of column 1 of table II to the person-years of table III it is found that the expected incidence of carcinoma in situ according to these rates agrees very well with the number observed. However, there were an additional six cases of early invasive cervical cancer produced by this population during the period of observation that were not expected if all cervical cancer begins as carcinoma in situ.

If we make the assumption for the moment that carcinoma in situ is not the origin of invasive cancer of the cervix, then the expected incidence of the invasive disease would be given by applying the age-specific rates of column 3 of table II to the person-years of table III. Under these conditions this population would have been expected to produce about 14 cases of beginning invasive cervical cancer. Actually six were observed. The deficit of 57 per cent could be taken to represent the proportion that ordinarily would have been supplied by carcinomas in situ progressing to invasion.

Again we might regard these six cases as having been carcinomas in situ at



TABLE III

OBSERVED AND EXPECTED CARCINOMA IN SITU AND INVASIVE CERVICAL CANCER INCIDENCE FROM RE-EXAMINATION OF WHITE WOMEN CYTOLOGICALLY NEGATIVE ON FIRST EXAMINATION —MEMPHIS AND SHELBY COUNTY, TENNESSEE [4]

Age Group	Person-years	Carcinoma in Situ		Invasive Cervical Cancer	
		Observed	Expected	Observed	Expected
20—	6,961	4	3.5	—	0.1
30—	8,661	9	6.1	2	3.5
40—	6,338	2	5.1	1	5.1
50—	2,849	5	2.8	2	3.4
60—	964	—	1.6	1	1.3
70—	194	—	0.3	—	0.3
Total	25,969	20	19.4	6	13.7

time of origin, with a short duration at this stage, and that they had become invasive by the time of discovery. The 26 cases of carcinoma in situ produced by the population while under observation then would have been 34 per cent in excess (26/19.4) of the number needed for progression to invasive cervical cancer. Considered from the standpoint of the deficit of invasive cancer, or the excess of carcinoma in situ, it appears that somewhere in the range of one-third to one-half of carcinomas in situ do not progress to invasive disease.

In table IV another set of data is shown giving the findings from cytologic re-examinations of women that were negative initially. Again the observed cases of carcinoma in situ and early invasive carcinoma are compared with the expected numbers computed with the rates of columns 1 and 3, respectively, of table II.

Here there is an even greater excess of carcinoma in situ incidence than is needed to satisfy the demands of invasive cancer. It is true that the rates for computing expected numbers come from another study involving another popula-

TABLE IV

OBSERVED AND EXPECTED CARCINOMAS IN SITU AND INVASIVE CERVICAL CANCER INCIDENCE FROM RE-EXAMINATION OF WHITE WOMEN CYTOLOGICALLY NEGATIVE ON FIRST EXAMINATION —SAN DIEGO [5]

Age Group	Person-years	Carcinoma in Situ		Invasive Cervical Cancer	
		Observed	Expected	Observed	Expected
20—	5,421	16	2.7	—	0.1
30—	7,339	11	5.1	—	2.9
40—	3,496	6	2.8	1	2.8
50—	1,280	1	1.3	1	1.5
60—	395	—	0.7	1	0.6
70—	82	—	0.0	—	0.1
Total	18,013	34	12.6	3	8.0

tion. It is unlikely, however, that if the necessary data were available for computing incidence rates of beginning invasive cervical cancer and carcinoma in situ for this population that such a difference would be overcome. Here again a substantial proportion of the invasive cancers appear as such, with a little less than two-thirds to be supplied from carcinomas in situ progressing to invasive disease.

## 5. Discussion and conclusions

I would not want to give the impression that the findings presented here are to be accepted as establishing any exact quantitative relationships between carcinoma in situ and invasive cancer of the cervix. The methodology proposed, however, offers the possibility of studying the interrelation of these lesions in a population. The data so far available are quite preliminary and much needs to be learned about the characteristics of the women that were given cytology examinations. The findings are worth noting in spite of their preliminary nature, as an indication of what is in need of further investigation.

The data from Memphis for invasive cervical cancer are probably a good approximation of the disease experience of the general population of that area. The estimate of three and a half years for the average duration of preclinical invasive cervical cancer over all the ages of women indicates there is considerable advantage in identifying these lesions by cytology rather than waiting for them to develop into clinically evident disease. The suggestion that this stage of the disease runs a more rapid course when developing in women in their 40's as compared to those younger and older is worth checking in other suitable data.

Computation of incidence rates for carcinoma in situ were made which would only be valid if all cervical cancer began as carcinoma in situ and all in situ carcinomas became invasive. The findings from periodic cytological re-examination of two groups of women who were cytologically negative initially throw considerable doubt on the validity of this assumption. In both groups about 40 per cent of the invasive cervical cancers that would have been produced by these populations if carcinomas in situ were not the source of the invasive disease did indeed appear as invasive disease. If carcinomas in situ needed only to make up the remaining 60 per cent of invasive disease, then there is excess incidence of carcinoma in situ in both populations. The excess in one population is considerably greater than the other. A possible explanation for the greater excess in the San Diego data may be that these patients were largely the clientele of obstetricians and gynecologists. Although the women had been found cytologically negative at least once, and many returned only for periodic cytology examinations, the possibility of bias for selecting those developing disease is great.

If the incidence of carcinoma in situ is in excess of that required to satisfy the needs of invasive cervical cancer, then the excess would need to terminate by regression. Further study of carcinomas in situ may provide a clue as to which have invasive potential.

## REFERENCES

- [1] G. N. PAPANICOLAOU, "New cancer diagnosis," *Proceedings of the Third Race Betterment Conference*, Battle Creek, Race Betterment Foundation, 1928.
- [2] G. N. PAPANICOLAOU and H. F. TRAUT, *Diagnosis of Uterine Cancer by the Vaginal Smear*, New York, The Commonwealth Fund, 1943.
- [3] J. E. DUNN, JR., "The relationship between carcinoma in situ and invasive cervical carcinoma. A consideration of the contributions to the problem to be made from general population data," *Cancer*, Vol. 6 (1953), pp. 873-886.
- [4] ———, "Preliminary findings of the Memphis-Shelby County uterine cancer study and their interpretation," *Am. J. Pub. Health*, Vol. 48 (1958), pp. 861-873.
- [5] J. E. DUNN, JR., T. A. SLATE, J. W. MERRITT, and P. L. MARTIN, "Findings for uterine cancer from one or more cytologic examinations of 33,750 women," *J. Nat. Cancer Inst.*, Vol. 23 (1959), pp. 507-527.
- [6] H. F. DORN and S. J. CUTLER, "Morbidity from cancer in the United States, Part I," *Public Health Monograph*, No. 29, U.S. Public Health Service, 1955.