

TREASURY DEPARTMENT
UNITED STATES PUBLIC HEALTH SERVICE
RUPERT BLUE, SURGEON GENERAL

INFLUENZA STUDIES

I. ON CERTAIN GENERAL STATISTICAL ASPECTS OF
THE 1918 EPIDEMIC IN AMERICAN CITIES

BY

RAYMOND PEARL, Ph. D.

*Professor of Biometry and Vital Statistics, School of Hygiene and Public
Health, Johns Hopkins University; Consultant in Vital
Statistics and Epidemiology
United States Public Health Service*

REPRINT No. 548

FROM THE

PUBLIC HEALTH REPORTS

AUGUST 8, 1919

(Pages 1743-1783)



19-27657

WASHINGTON
GOVERNMENT PRINTING OFFICE
1919

Monograph.

RC 150
5
A15PX
no. 1

No. of J.
NOV 26 1919



9

6/16/1919

INFLUENZA STUDIES.

I. ON CERTAIN GENERAL STATISTICAL ASPECTS OF THE 1918 EPIDEMIC IN AMERICAN CITIES.¹

By RAYMOND PEARL, Ph. D., Professor of Biometry and Vital Statistics, School of Hygiene and Public Health, Johns Hopkins University; Consultant in Vital Statistics and Epidemiology, United States Public Health Service.

I. Introduction.

The pandemic of influenza which swept over the world in 1918 was the most severe outbreak of this disease which has ever been known, and it takes an unpleasantly high rank in the roster of epidemics generally. It is certainly impossible now, and perhaps always will be, to make any precise statement of the number of people who lost their lives because of this epidemic. But it is certain that the total is an appalling one. Undoubtedly a great many more people died from this cause than from all causes directly connected with the military operations of the Great War. In the United States alone conservative estimates place the deaths from the influenza epidemic at not less than 550,000, which is approximately five times the number (111,179) of American soldiers officially stated² to have lost their lives from all causes in the war. And the end of the epidemic is by no means yet reached. In England and Wales the curve of mortality from influenza was even in 1907, seventeen years after the epidemic of 1890, higher than it was in any of the 40 years preceding 1890. The decline in the mortality rate after the 1848 epidemic in Great Britain was similarly slow.³ There is no evident reason to suppose that conditions following the first explosion of this present epidemic will be essentially different from those which obtained in the earlier cases.

For two reasons the hygienist and epidemiologist should be interested in the intensive study, from every possible angle, of the present pandemic. In the first place, owing to the advances which have been made in every branch of medical science since the epidemic of 1890, there is now available a much more adequate investi-

¹ Papers from the Department of Biometry and Vital Statistics, School of Hygiene and Public Health, Johns Hopkins University, No. 5. This investigation was carried on in consultation with the United States Public Health Service, Office of Field Investigations on Influenza, Dr. W. H. Frost, surgeon in charge. Reprint from the Public Health Reports, vol. 34, No. 32, Aug. 8, 1919, pp. 1743-1783.

² As of date Apr. 30, 1919.

³ Cf. Article on "Influenza" in Encyclopedia Britannica, 11th Edition, for a conveniently accessible verification of these statements.

gational armament with which to attack the problems raised by such an epidemic than was the case earlier. Furthermore, the whole machinery for getting accurate records of the incidence and results of the outbreak are much better now than they were 30 years ago. This is particularly true in the United States. The records of mortality connected with the present epidemic are unquestionably more complete and accurate than any that have ever before been available in this country for any epidemic of anything like so great extent or force.

In the second place, the very magnitude of this epidemic is in itself a challenge to the whole medical profession. The hygienists of the world are the standing army, which is, in theory at least, maintained by society to organize and hold the defenses against such dread invaders as these. Such a blow as the present one may well inspire a slogan like that which saved Verdun, "*Ils ne passeront pas.*" If every epidemiologist does not take advantage of the present opportunity to investigate with all possible thoroughness epidemic influenza, to the end of making a better defense next time, he will have been derelict in his plain duty.

• The present paper is intended as a first contribution toward the statistical analysis of certain phases of the 1918 influenza epidemic. It will be followed by further papers in the same series dealing with other aspects of the problem. In the first studies in the series attention will be confined entirely to the *mortality* records of some forty of the larger cities of the United States. The reason for this limitation to mortality only and to large cities is that accurate and reliable data within these limitations are now available, and the same can not be said of morbidity records, on anything like so general a scale. Later it is expected that sufficiently accurate and extensive morbidity statistics of the epidemic to warrant statistical analysis will be available.

The data of this study are taken primarily from the Weekly Health Index.¹ On account of varying medical opinions as to the properly reportable terminal cause of death of persons dying after having had influenza during this epidemic, it has been thought safest to use death rates from all causes for study, rather than those specifically reported to the registrar as due to influenza or pneumonia. Consequently, we shall deal with death rates from all causes in discussing the present epidemic. This makes no practical difference in the statistical results, because the deviation of the curves of total mortality from their normal course during the epidemic was due entirely to causes inherently associated with the epidemic itself. The use of the death rate from all causes during the epidemic has the fur-

¹ A typewritten publication issued weekly by the Bureau of the Census, and compiled under the direction of Dr. W. H. Davis, Chief for Vital Statistics.

ther advantage that it takes into account those deaths which occur from diseases of the heart or kidneys some weeks or months after an attack of influenza from which the patient has apparently recovered, but which in reality are responsible for the fatal break-down of a part of the organic machinery which had long been weak, and only required for its complete collapse some such strain as the attack of influenza superimposed.

The general problem with which the first study in this series will have to do is that of the statistical analysis of the *first explosive outbreak* of epidemic mortality in large American cities. As will presently appear, there was an extraordinary degree of variation amongst the different cities in respect of the initial force and duration of this first explosion. These *differences* between cities in respect of the severity and suddenness with which they were attacked by the disease constitute the first great problem which the epidemic has raised. What factors had a causal influence in determining this great observed variation among cities? The full significance of this problem will be apparent when the facts of variation in force of explosive outbreak are before us. The first task of this study is to present the data in such a manner as to bring out the real extent and magnitude of the variation in the epidemic.

I am indebted to Mr. John Rice Miner for the greater portion of the laborious arithmetic connected with this investigation.

II. General Survey of the Mortality Curves.

In order to get in hand the general problem it is desirable to examine with some care the mortality by weeks in each of the cities dealt with. To this end Figures 1 to 6 have been prepared. On these diagrams are plotted, for each city, the annual death rates per 1,000 population from all causes, for each week, the data being those of the Weekly Health Index. The plotting is done on a logarithmic scale of ordinates (rates) and an arithmetic scale of abscissæ (weeks).¹ The curves begin with the week ended July 6, 1918, and continue to 1919. The scale is the same for all diagrams, though different combinations of parts of the logarithmic "decks" are used in certain cases in order to fit the diagrams to the page.

Anyone examining these curves thus collected together on a uniform scale for comparison can not fail to be impressed by the fact that there is an extraordinary amount of difference between different cities in respect of the force with which they were struck by the epidemic at its initial outbreak. Compare, for example, the Albany, Boston, Baltimore, Dayton, or Philadelphia curves with those for Atlanta, Indianapolis, Grand Rapids, Milwaukee, or Minneapolis. The former curves show an initial sudden explosive outbreak of great

¹ For a discussion of the advantages of "arithlog" paper see Fisher, I. "The 'Ratio' Chart for plotting Statistics." Quarterly Publications Amer. Stat. Assoc., 1917, pp. 577-601.

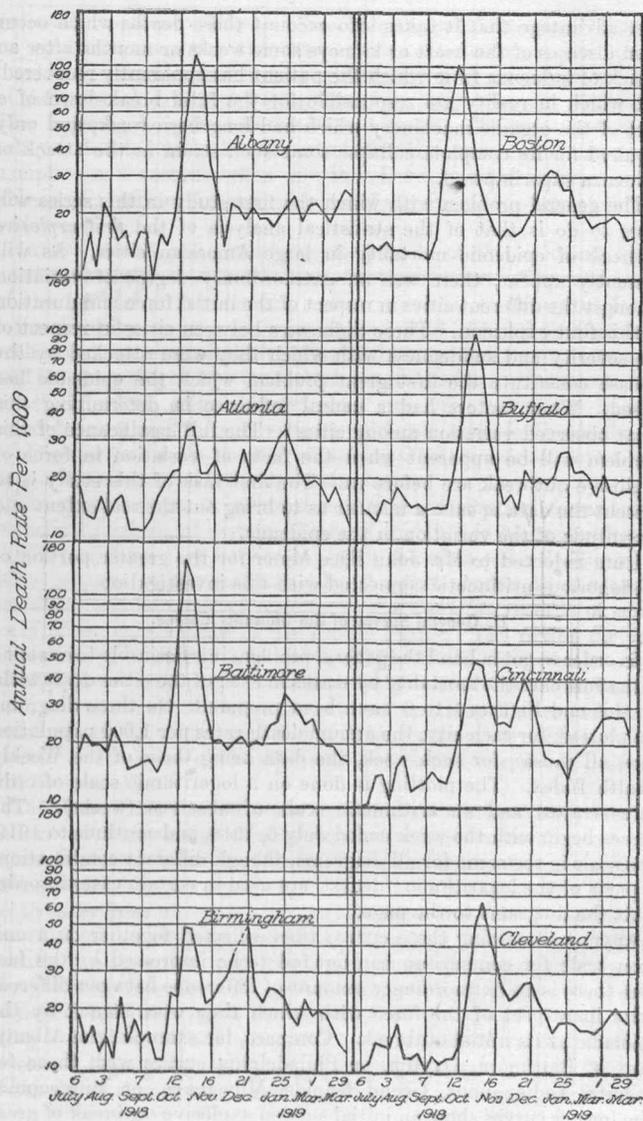


FIGURE 1.—Annual death rates, by weeks, per 1,000 population, for 8 cities.

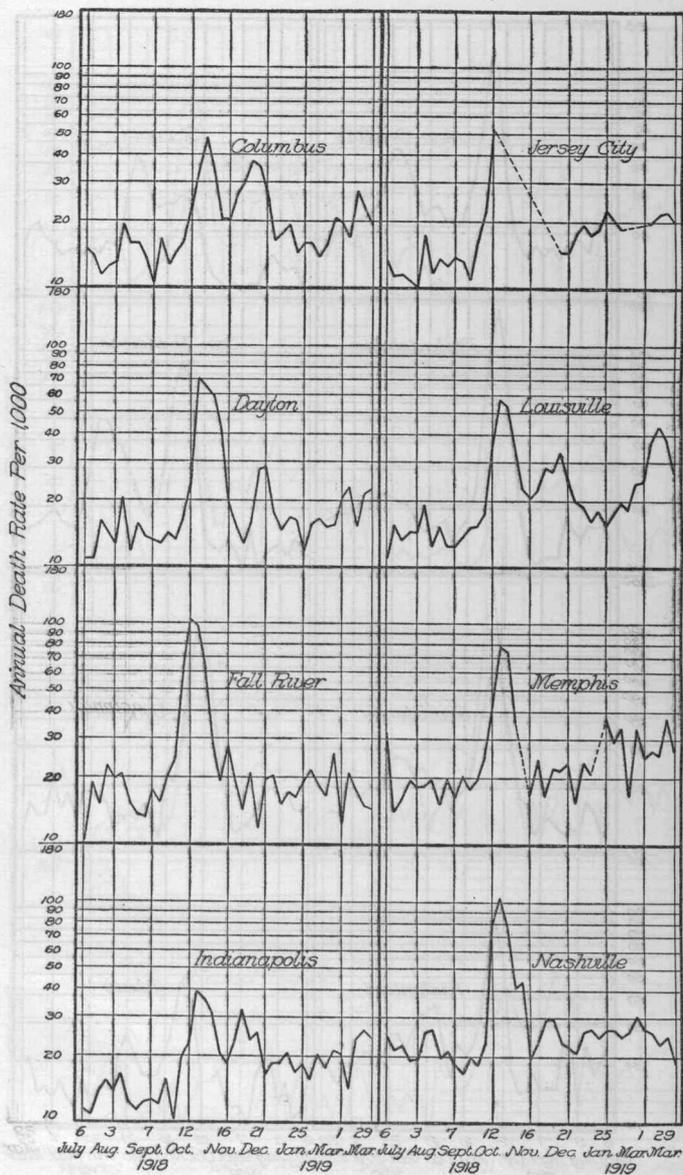
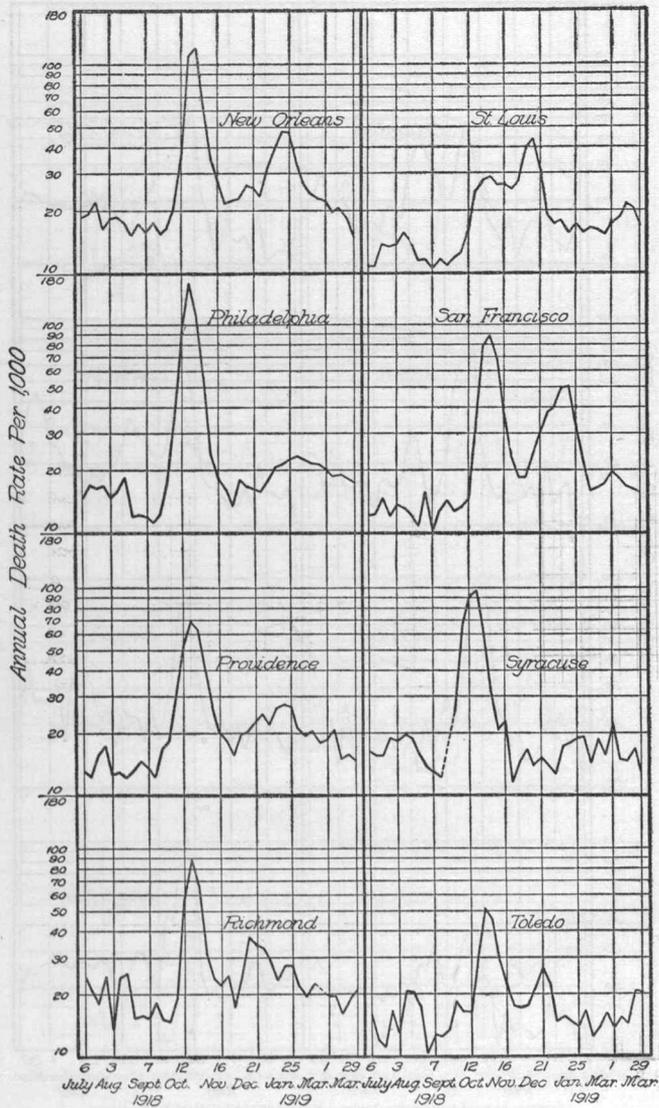


FIG. 2.—Annual death rates, by weeks, per 1,000 population, for 8 cities.



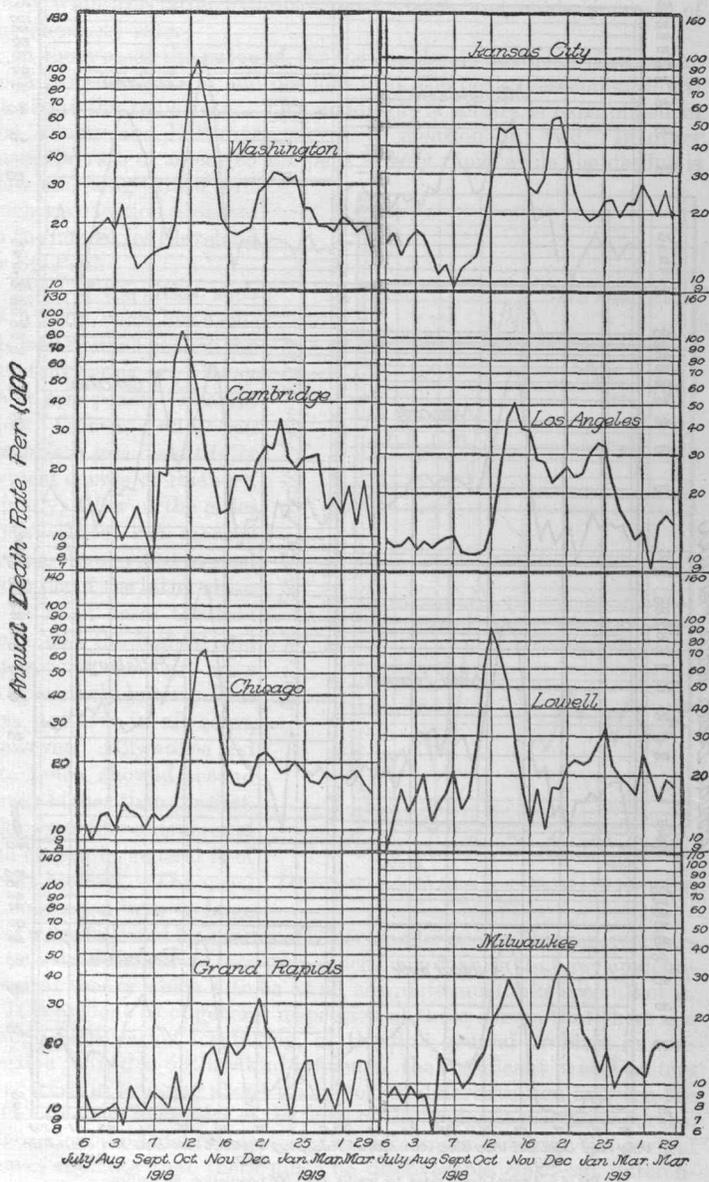


FIG. 4.—Annual death rates, by weeks, per 1,000 population, for 8 cities.

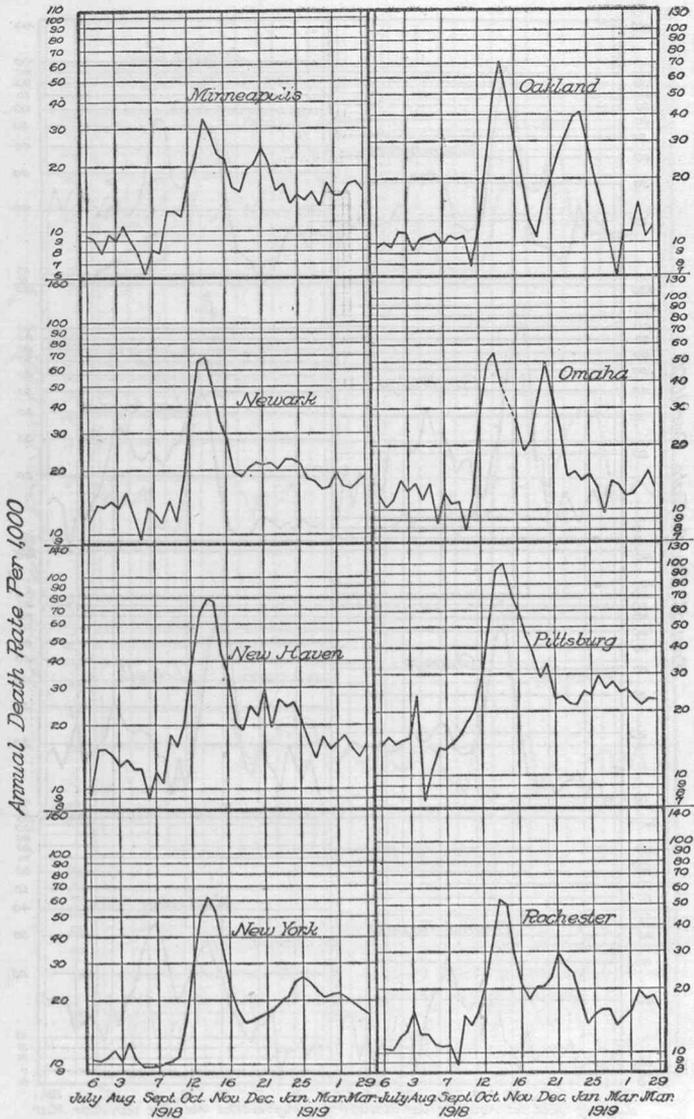


FIG. 5.—Annual death rates, by weeks, per 1,000 population, for 8 cities.

force, while the latter exhibit a much slower and milder increase of the mortality rate.

In some cases the curve of the first epidemic outbreak rises to the peak (ascending limb) and declines from the peak (descending limb) at about the same rate. This condition of affairs is exemplified in the Albany and Baltimore curves, to mention but two. In other cases the rate of ascent to the peak is very rapid while the decline is slow and long drawn out. Such a condition is shown in the curves for Cleveland or St. Paul.

Some of the cities, such as Albany, show but a single well-defined peak in the mortality curve. Many show two peaks. Boston, New Orleans, and San Francisco give beautifully typical curves of this sort. Finally, a few of the cities show three well-marked peaks. Louisville is a good example of the latter class.

In most cases the first peak was the highest and the second and third were progressively lower. This was not true in all cases, however. Milwaukee and St. Louis showed second peaks higher than the first. The wave-like character of the curves in general is of great interest. The usual phenomenon was a large first wave followed by a series of other smaller ones. This general characteristic of the curves is so pronounced and definite that any epidemiological theory which is to be at all adequate must take account of it.

It is evident from general inspection of these curves that there is a strong justification for taking, as the first general problem in connection with this outbreak of influenza, the significant causal factors concerned in bringing about this observed differentiation between the different cities in respect of the form of the epidemic mortality curves. The extent and definiteness of the differences between the several curves indicate that there must be discoverable clean-cut differentiating factors which influenced the influenza death rates.

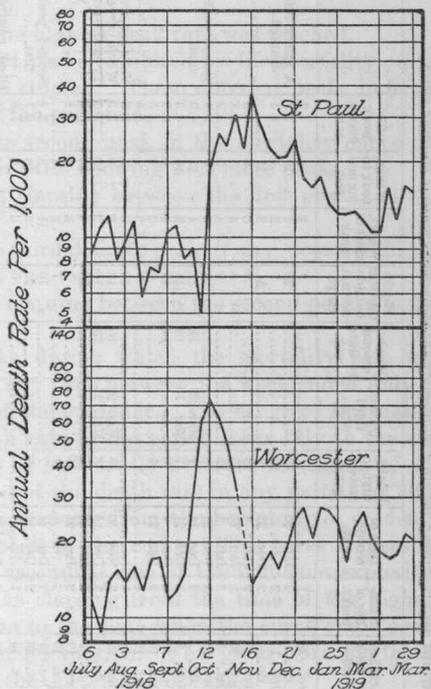


FIG. 6.—Annual death rates, by weeks, per 1,000 population, for 2 cities.

TABLE I.—Certain data regarding the time relation of the influenza epidemic in large cities.

Name of city.	Highest peak rate.	Reached on week ended—	Number of peaks.	Date of second peak.	Weeks, first to second peak.	Date of third peak.	Weeks, second to third peak.	Weeks rate was outside normal range.	Weeks, start to peak.	Weeks, peak to normal rate.	25-week excess rate.	Notes.
Albany, N. Y.	112.0	Oct. 26	1					7	3	4	4.7	A minor intermediate third peak not counted.
Atlanta, Ga.	35.1	do.	2	Jan. 25	13			23	4	19	2.7	
Baltimore, Md.	148.3	Oct. 19	2	do.	14			7	3	4	6.1	
Birmingham, Ala.	48.6	do.	2	Dec. 21	9			14	3	11		Second peak small as compared with first.
Boston, Mass.	100.2	Oct. 5	2	Jan. 4	13			9	3	6	6.5	
Buffalo, N. Y.	99.7	Oct. 26	2	Jan. 25	13			8	4	4	5.8	
Cambridge, Mass.	83.0	Oct. 5	2	Jan. 4	13			9	4	5	5.9	Second peak small as compared with first.
Chicago, Ill.	64.2	Oct. 26	2	Dec. 21	8			8	4	4	3.8	
Cincinnati, Ohio	49.7	do.	3	Dec. 14	7	Mar. 22	14	17	4	13	4.0	
Cleveland, Ohio.	65.6	Nov. 2	3	Dec. 21	7	Mar. 15	12	13	3	10	4.0	Very low, irregular curve. Whole curve low. Typical 3-peak curve.
Columbus, Ohio	47.2	Oct. 26	3	Dec. 7	6	do.	14	12	3	9	3.2	
Dayton, Ohio	71.6	Oct. 19	2	Dec. 21	9			8	2	6	3.5	
Fall River, Mass.	100.4	Oct. 12	1					7	3	4	5.8	Very low, irregular curve. Whole curve low. Typical 3-peak curve.
Grand Rapids, Mich.	31.6	Dec. 14	1					12	8	4	1.5	
Indianapolis, Ind.	39.0	Oct. 19	3	Nov. 30	6	Mar. 22	16	12	3	9	2.5	
Louisville, Ky.	57.6	do.	3	Dec. 14	8	Mar. 15	13	14	2	12	3.6	Second peak definitely higher than first.
Los Angeles, Calif.	51.9	Nov. 2	2	Jan. 18	11			20	4	16	5.2	
Lowell, Mass.	89.8	Oct. 12	2	Jan. 25	15			7	3	4	5.1	
Memphis, Tenn.	73.3	Oct. 19	2	do.	14			7	3	4	5	Second peak definitely higher than first.
Milwaukee, Wis.	34.4	Dec. 14 ¹	2	Dec. 14	7			16	11	5	2.9	
Minneapolis, Minn.	33.1	Oct. 26	2	Dec. 21	8			22	6	16	2.7	
Nashville, Tenn.	104.2	Oct. 19	1					6	2	4	7.8	Both possibly 1-peaked. But 2 seems on whole best, especially for New Haven. Beautiful typical 2-peak curve.
Newark, N. J.	69.0	Oct. 26	2	Jan. 18	12			20	4	16	5.1	
New Haven, Conn.	78.8	do.	2	Dec. 21	8			21	6	15	5.6	
New Orleans, La.	120.1	do.	2	Jan. 18	12			7	3	4	7.2	It is possible that this is 3-peaked.
New York, N. Y.	61.6	do.	2	Jan. 25	13			9	4	5	4.7	
Oakland, Calif.	79.5	Nov. 2	2	Jan. 18	11			18	3	15	5.9	
Omaha, Nebr.	65.1	Oct. 26	2	Dec. 14	7			15	3	12	7.3	Second peak low.
Philadelphia, Pa.	158.3	Oct. 19	2	Feb. 1	15			7	3	4	7.3	
Pittsburgh, Pa.	100.7	Nov. 2	2	do.	13			11	4	7	8.0	
Providence, R. I.	69.8	Oct. 19	2	Jan. 18	13			10	4	6	5.3	Possibly this should be 2.
Richmond, Va.	88.4	do.	2	Dec. 14	8			5	2	3	5	
Rochester, N. Y.	85.8	Oct. 26	3	Dec. 21	8			16	4	12	2.7	
St. Louis, Mo.	44.0	Dec. 14 ²	3	Dec. 14	6	Mar. 15	13	15	10	5	3.0	Abnormal curve. Typical 2-peak.
St. Paul, Minn.	36.2	Nov. 16	1					19	6	13	3.3	
San Francisco, Calif.	89.0	Nov. 2	2	Jan. 18	11			17	3	14	7.5	
Syracuse, N. Y.	96.7	Oct. 19	3	Dec. 21	8	Mar. 8	11	16	4	12	2.1	Typical 2-peak.
Toledo, Ohio.	51.1	Oct. 26	3	Dec. 28	9	Mar. 22	12	6	2	4	2.1	
Washington, D. C.	109.3	Oct. 19	2	do.	10			6	3	3	6.6	
Worcester, Mass.	72.2	Oct. 12	2	do.	11			7	3	3	4	

¹ First peak date Oct. 26.² First peak date Nov. 2.

III. Classification of the Data.

As a first step in the analysis it is desirable to make certain rough classifications of the facts brought out by the mortality curves. To this end Table I has been prepared. In this table are set forth the following data regarding each of the cities:

1. The highest peak death rate attained in any week of the epidemic up to March 29, 1919.
2. The date¹ on which the highest peak rate was reached.
3. The number of distinct peaks exhibited by the mortality curve within the time period here studied. These different peaks indicate recrudescences or waves of the epidemic.
4. The date at which the second peak in the mortality curve occurred, in the case of those cities showing 2 or more peaks.
5. The number of weeks elapsing between the first peak and the second.
6. The date at which the third definite peak, if any, occurred in the mortality curve.
7. The number of weeks elapsing between the second peak and the third.
8. The number of weeks during which the mortality rate was higher than it had been at any time between the week ended July 6, 1918, and the beginning of the epidemic. The range of fluctuation of the weekly annual death rate in the period from July to the end of September was held to be sufficiently accurate indication of the normal range of fluctuation of the death rate in any particular city.
9. The number of weeks elapsing from the beginning of epidemic mortality to the highest peak of the curve. This gives a measure of the time factor on the ascending side of the epidemic explosion.
10. The number of weeks elapsing from the time of the highest peak of the mortality curve to the time when the curve came again within the normal range of fluctuation. This gives the time factor on the descending limb of the epidemic outbreak.
11. The excess mortality rate, over the normal for the same season of the year for the same places, for the 25 weeks between September 8, 1918, and March 1, 1919. These figures were issued as a supplement to the Weekly Health Index by the Census Bureau.²

From this table a number of points present themselves for discussion. They may best be taken up in separate sections, in order of the successive rubrics of the table.

1. *Maximum peak death rates.*—The highest or maximum peak rate of mortality during the epidemic varied greatly, having ranged from

¹ It is to be understood that all dates here and throughout are as of "weeks ended" on the specified date. The original statistics are given only in weeks and hence any finer time differentiation is impossible.

² Cf. Public Health Reports, vol. 34, No. 11, p. 505, 1919.

31.6 in the case of Grand Rapids, Mich., to 158.3 in the case of Philadelphia.

The distribution of the different maximum peak rates over this range is shown in detail in Table II.

TABLE II.—Showing the frequency of occurrence of different maximum peak death rates during the epidemic.

Maximum peak rates.	Number of cities.
30.0-39.9.....	6
40.0-49.9.....	4
50.0-59.9.....	5
60.0-69.9.....	5
70.0-79.9.....	5
80.0-89.9.....	4
90.0-99.9.....	2
100.0-109.9.....	5
110.0-119.9.....	1
120.0-129.9.....	1
130.0-139.9.....	0
140.0-149.9.....	1
150.0-159.9.....	1
Total.....	40

From Table II it appears that in the 40 cities considered the peak rates which were of the most frequent occurrence were, generally speaking, rates below 70. Twenty out of the 40 fell below that figure. Only 9 out of the 40 cities showed a maximum peak rate of 100 or more. Up to a maximum peak rate of 70 the distribution is very even in the four classes of 10 points each in the rate. From 70 on it falls off rapidly, with the single exception of the class of rate from 100 to 109.9, which has a frequency of 5.

The detailed distribution of the maximum peak rate is shown graphically in Figure 7.

TABLE III.—Constants for maximum peak death rates.

Constant.	Value.
Mean maximum peak rate.....	73.9±3.2
Median maximum peak rate.....	70.0±4.0
Standard deviation.....	30.3±2.3

Three of the cities, Milwaukee, Kansas City, and St. Louis, show higher maximum peak rates on the second wave than on the first.

2. *Date of occurrence of maximum peak rate.*—The date of the week in which the maximum peak rate occurred is given in the third column of Table I. It will be seen that the earliest date, October 5, occurs but twice, namely, in Boston and Cambridge. These two cities, of course, are in a demographic sense practically a single unit though politically separate. At the other extreme the latest maximum peak rate date is December 14. The cities showing a culmina-

tion of the epidemic mortality during the week which ended on this latter date are Grand Rapids, Milwaukee, and St. Louis. Grand Rapids has an extremely peculiar curve, unlike that of any other city in the country. Milwaukee and St. Louis are two of the cities showing the second peak higher than the first, so in these two cases the date in the third column of Table I refers to the second peak, while in all other cities it refers to the first peak. On these accounts the upper range end for maximum peak date should probably not

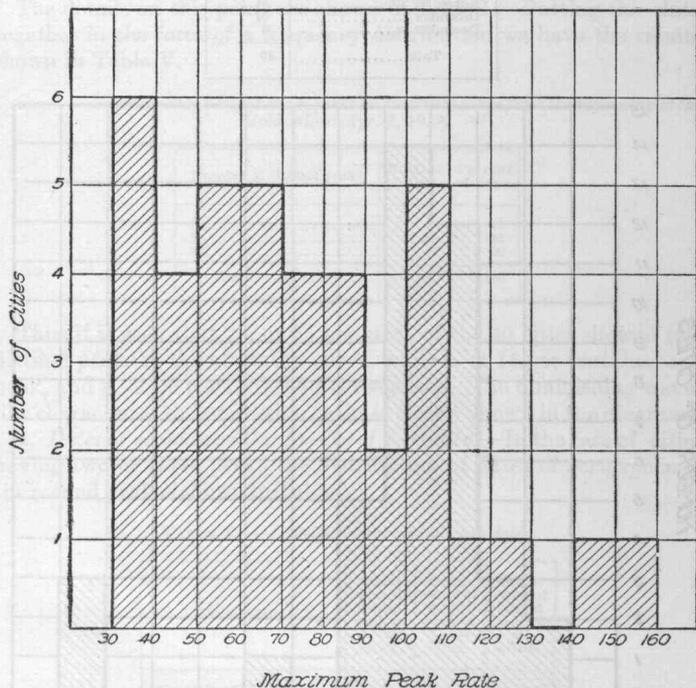


Fig. 7.—Distribution of maximum peak death rates in 40 cities. Certain constants of the distribution shown in Table II are exhibited in Table III.

be taken as December 14, but as November 2, since the only other later date, November 16, appears in a single case, St. Paul, and the curve for that city is again abnormal. There are five cities showing the peak of the mortality curve in the week ended November 2, namely, Cleveland, Los Angeles, Oakland, Pittsburgh, and San Francisco.

The distribution of maximum peak dates is shown in Table IV, and graphically in Figure 8.

TABLE IV.—Distribution of dates of maximum peak mortality.

Maximum peak in week ended—	Number of cities.
October 5.....	2
October 12.....	3
October 19.....	12
October 26.....	14
November 2.....	5
November 9.....	0
November 16.....	1
November 23.....	0
November 30.....	0
December 7.....	0
December 14.....	3
Total.....	40

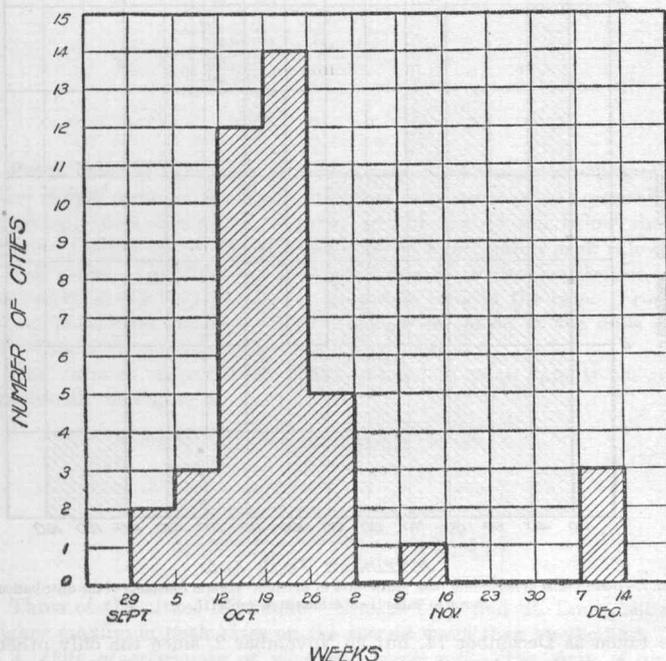


FIG. 8.—Distribution of peak dates of the epidemic.

Using all the data, we find the following constants for date of maximum peak.

Mean peak date = October 23 \pm 1.68 days.

Standard deviation in peak date = 15.75 \pm 1.19 days.

These constants will serve as a useful record of the time factor in the epidemic of the autumn of 1918 in American cities.

Thirty-one out of the 40 cities had attained the peak rate of mortality prior to November 2.

3. *Number of peaks in mortality curve.*—It is clear from the diagrams already shown that there was considerable variation in the different cities in respect of the number of epidemic mortality peaks exhibited.

The details on this point are shown in Table I. Putting the data together in the form of a frequency distribution we have the results shown in Table V.

TABLE V.—Showing number of distinct peaks in mortality curve from the beginning of the epidemic to Apr. 1, 1919.

Number of distinct peaks.	Number of cities.	Per cent of cities.
1.....	6	15
2.....	26	65
3.....	8	20
Total.....	40	100

Thus it is seen that 26, or 65 per cent, of the 40 cities showed two distinct peaks in the mortality curve, while 6, or 15 per cent, had one peak, and 8, or 20 per cent, had three peaks. The diminishing wave-like character of the successive peaks is clearly shown in the diagrams.

4. *Dates of second and third peaks of mortality.*—In the case of cities having two or three peaks the distribution of dates of occurrence of the second peak is shown in Table VI.

TABLE VI.—Distribution of second-peak dates.

Week ended—	Occurrence of second peak in 2-peak cities.	Occurrence of second peak in 3-peak cities.	Occurrence of second peak in all cities.
November 30.....		1	1
December 7.....		1	1
December 14.....	3	3	6
December 21.....	5	2	7
December 28.....	2	1	3
January 4.....	2		2
January 11.....			
January 18.....	6		6
January 25.....	6		6
February 1.....	2		2
Total.....	26	8	34

Certain interesting facts stand out clearly from this table. In the 8 cities which had three distinct peaks of mortality the second peak came early—prior to December 28. The distribution for the 26

cities having two peaks of mortality is distinctly bimodal, 12 of them showing a mode for the week ended December 21, and 14 a mode somewhere in the weeks of January 18 and 25. No city had a second peak of mortality in the week ended January 11.

Table VII gives the distribution of dates of the third peak of mortality.

TABLE VII.—*Distribution of third peak dates.*

Week ended—	Occurrence of third peak.
March 8.....	1
March 15.....	4
March 22.....	3
Total.....	8

Here the observed mode evidently falls somewhere in the week ended March 15.

The data of Tables VI and VII are shown graphically in Figure 9.

The figures and diagram at once suggest that the group of 12 two-peak cities showing the second peak somewhere between December 7 and January 4 were cities which at that time were presumably destined to show a third distinct wave and peak of mortality, but in which for some reason not now apparent the third wave did not eventuate. In contradistinction to these stand the 14 cities showing a second peak of mortality between January 11 and January 21. These latter are presumably cities in which the complex of factors determining the form of the mortality curve was such as to lead definitely to a two, and only two, peaked curve. This idea will be substantiated by further evidence to be presented immediately.

As a matter of record of the epidemic in American cities, the mean dates calculated from Tables VI and VII are given in Table VIII.

TABLE VIII.—*Constants for dates of second and third mortality peaks.*

Item.	Mean.	Standard deviation.
Date of second peak.....	Jan. 1 ± 2.13 days	18.40 ± 1.51 days.
Days from beginning of October to second peak.....	92.26 days.....	
Date of third peak.....	Mar. 14 ± 1.10 days.	4.63 ± 0.78 days.
Days from beginning of October to third peak.....	165.25 days.....	

Putting all the data together we find for the whole group of cities the following *average* relations:

(a) Days from average date to maximum peak in all cities to second peak in cities showing two or three mortality peaks = 69.26.

(b) Days from date of second peak, in all cities showing two or more peaks, to third peak, in cities having three mortality peaks = 72.99.

These relations seem at first sight to point to a cycle of about 10 weeks' duration in the secondary mortality waves of this influenza

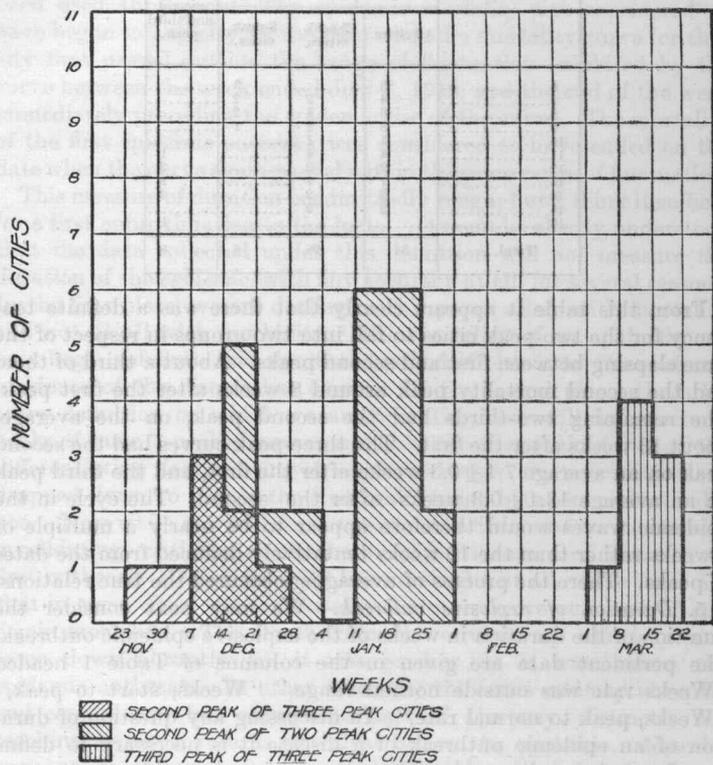


FIG. 9.—Frequency of occurrence of second and third peaks of mortality at different dates.

epidemic, after the first wave. This point can, however, be more accurately discussed by reference to the data set forth in Table I on the number of weeks elapsing between the successive peaks.

These data are presented in the form of frequency distributions in Table IX.

TABLE IX.—Frequency distributions of number of weeks elapsing between successive mortality peaks.

Number of weeks.	Number of cities.			
	Between first and second peak.			Between second and third peak.
	All cities.	2-peak cities.	3-peak cities.	
6.....	3	3
7.....	4	2	2
8.....	6	4	2
9.....	3	2	1
10.....	1	1
11.....	4	4	1
12.....	2	2	2
13.....	7	7	2
14.....	2	2	2
15.....	2	2
16.....	1
Total.....	34	26	8	8

From this table it appears clearly that there was a definite tendency for the two-peak cities to fall into two groups in respect of the time elapsing between first and second peaks. About a third of them had the second mortality peak around 8 weeks after the first peak. The remaining two-thirds had the second peak, on the average, about 13 weeks after the first. The three-peak curves had the second peak on an average 7.1 ± 0.3 weeks after the first, and the third peak on an average 13.1 ± 0.3 weeks after the second. The cycle in the epidemic waves would therefore appear to be nearly a multiple of 7 weeks rather than the 10 weeks tentatively deduced from the dates of peaks. There the process of averaging obscured the true relations.

5. *Duration of explosive outbreak.*—We may next consider the question of the duration in weeks of the explosive epidemic outbreak. The pertinent data are given in the columns of Table 1 headed "Weeks rate was outside normal range," "Weeks, start to peak," "Weeks, peak to normal rate." In discussing any question of duration of an epidemic outbreak of a disease it is necessary to define sharply and usually arbitrarily what are to be taken as limiting points. It is always difficult, and usually impossible, to define these limiting points precisely and logically so that no one will or can criticize their location. The point has recently been discussed by Hitchcock and Carey¹ who say: "The difficulty * * * lies in deciding at just what point an undue prevalence or outbreak becomes epidemic." The general epistemological principle to be observed is clearly this: That since it is usually impossible to say with mathematical precision, in the case of an endemic disease, exactly when an epidemic outbreak begins or ends one must, in order to avoid

¹ Hitchcock, J. S. and Carey, B. W., "A Median Epidemic Index. Amer. Jour. Public Health, Vol. IX, pp. 355-357. 1919.

unconscious bias in dealing with a series of different localities, lay down an arbitrary rule and follow it absolutely. Then the results will be correct *relative to each other*, even though there may be room for argument as to whether they are absolutely correct or not. Following this principle the following rule was laid down and has been used throughout: The epidemic *mortality* was considered to have begun in any city on the date when the mortality curve for that city first passed outside the range of fluctuation exhibited by the curve between the week ended July 6, 1918, and the end of the week immediately preceding the epidemic rise of the curve. The mortality of the first epidemic outbreak was considered to have ended on the date when the curve again passed within the same range of fluctuation.

This measure of duration is admittedly rough, but I think it suffices for a first approximation to the facts. It must be clearly understood that the data collected under this definition will not measure the duration of the *epidemic*, with any accuracy at all, for several reasons. In the first place, we are dealing in this paper solely with mortality and not at all with morbidity. The mortality of an epidemic can only begin a definite and significant period of time *after* the epidemic incidence of the disease has begun. In the second place, the arbitrary definition on which we are operating here will include both peaks of some 2-peaked curves and only the first peak of others, the differentiating factor being of course whether the mortality curve dropped down to within the "normal" range between peaks or did not. Now while this will seem to some a serious, not to say totally invalidating, criticism of the here defined measure of duration of first outbreak, I think it really has no weight at all. The facts are that in some cities (A) there was a sharp explosive outbreak of epidemic mortality. The death rate curve went up abruptly and came down abruptly till it was as low as it was before the epidemic outbreak. In other cities (B) the curve went up abruptly and came down, but only some part of the way, distinctly not reaching so low a rate as prevailed before the epidemic. Now by any canons of common sense it would seem clear that in the A cities the *particular epidemic outbreak* about which we are talking came to an end when the death rate was again normal for the locality and season. Subsequently the death rate may have again risen abruptly. But if it did it was a *new* and distinct epidemic outbreak, temporally and spatially related to the first outbreak if one likes, but definitely separated from it by a longer or a shorter period in which the mortality rate was *normal*. Conversely in the B cities even though the mortality rate did decline from the maximum peak rate, still it did not go back to normal, or in other words it remained an *epidemic mortality*, in the common sense of that word. The rate after this depression may have risen to a new second peak,

but all the time it was part of the same epidemic outbreak. Thus it clearly appears that there is a real distinction between the A cities and the B cities. This distinction is reflected perfectly in the duration definition here adopted, and would be wholly lost in any scheme of measuring duration by peaks alone. It only needs to be kept firmly fixed in mind that we are here measuring the length of time during which the death rate was higher than the normal death rate for the same city, in the first continuous outbreak of influenza mortality.

We may first consider the total number of weeks that the mortality was outside the July to September range of fluctuation. The frequency distribution is given in Table X.

TABLE X.—*Frequency distribution of cities in respect of number of weeks mortality curve was outside "normal" range of fluctuation in first outbreak.*

Weeks.	Number of cities.
5.....	11
6.....	3
7.....	8
8.....	3
9.....	4
10.....	1
11.....	1
12.....	3
13.....	1
14.....	2
15.....	2
16.....	2
17.....	2
18.....	1
19.....	1
20.....	2
21.....	1
22.....	1
23.....	1
Total.....	40

The range of variation in the duration of the first outbreak of epidemic mortality, as here defined, is great, from five weeks on the one hand (Richmond, Va.) to 23 weeks on the other (Atlanta, Ga.). So great is this variation that its general trend is not easily comprehended until the figures are somewhat combined. If that is done, certain general relations appear. First of all, it is to be noted that 20 cities, exactly one-half the total number, showed a duration as here defined of 10 weeks or less, while in the other half the duration was 11 weeks or over. The median duration was then 10.5 weeks.

In general, the tendency was for the shorter duration to occur more frequently. This is well shown by Figure 10, which is plotted from the last column of combined figures in Table X.

Considerably the largest single area in the histogram is the first one covering durations of five to eight weeks inclusive. The frequencies for the longer periods, shown in four-week groups, become successively smaller.

From the ungrouped data of Table X the following constants have been calculated:

Mean duration of epidemic mortality in the first outbreak = 11.90 ± 0.55 weeks.

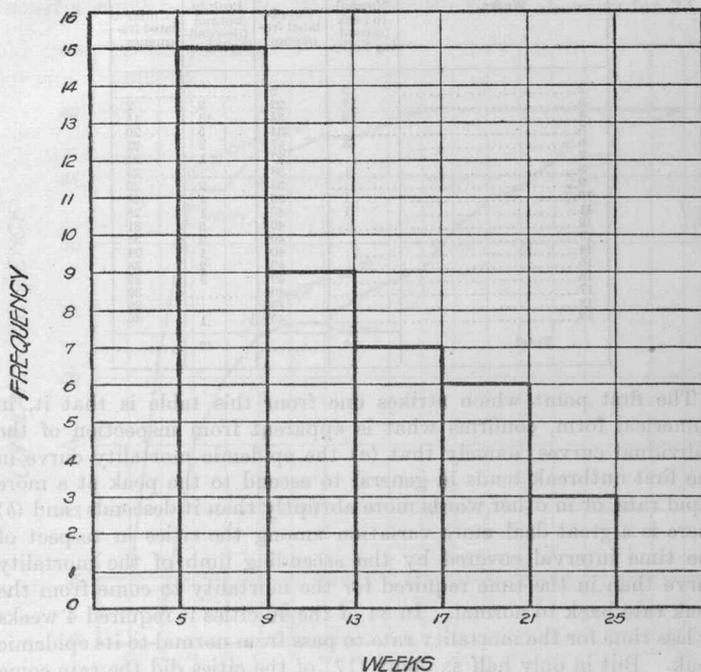


FIG. 10.—Frequency of different durations of the first outbreak of epidemic mortality.

Standard deviation = 5.17 ± 0.39 weeks.

We may next consider the two limbs of the explosive mortality curve. The frequency distributions for the time duration of the ascending limbs and the descending limbs are given in Table XI.

TABLE XI.—Frequency distributions for two moieties of epidemic mortality curve (first outbreak).

Weeks.	Frequency.			
	Normal to peak (ascending limb).	Cumulated frequency.	Peak to normal (descending limb).	Cumulated frequency.
2	5	5		
3	17	22	2	2
4	12	34	13	15
5		34	5	20
6	3	37	3	23
7		37	1	24
8	1	38		24
9		38	2	26
10	1	39	1	27
11	1	40	1	28
12		40	3	31
13		40	2	33
14		40	1	34
15		40	2	36
16		40	3	39
17		40		39
18		40		39
19		40	1	40
Total	40		40	

The first point which strikes one from this table is that it, in numerical form, confirms what is apparent from inspection of the individual curves, namely that (a) the epidemic mortality curve in the first outbreak tends in general to ascend to the peak at a more rapid rate, or in other words more abruptly than it descends; and (b) there is a great deal more variation among the cities in respect of the time interval covered by the ascending limb of the mortality curve than in the time required for the mortality to come from the peak rate back to normal. In 34 of the 40 cities it required 4 weeks or less time for the mortality rate to pass from normal to its epidemic peak. But in only half as many (17) of the cities did the rate come down from its peak to normal again in a period of 4 weeks or less.

The constants of the two distributions are as follows:

Mean time from normal mortality rate to peak = 3.90 ± 0.21 weeks.

Standard deviation in time from normal mortality rate to peak = 1.93 ± 0.15 weeks.

Mean time from peak mortality rate to normal = 8.00 ± 0.50 weeks.

Standard deviation in time from peak mortality rate to normal = 4.68 ± 0.35 weeks.

From these figures it appears that on the average it took about twice as many weeks for the mortality curve to come back from its peak condition to the normal again, as were required for the increase from normal to peak at the beginning of the explosion. In round figures, the ascending limb of the mortality curve occupied about a month and the descending limb about two months.

The differences between the two distributions of Table XI are well shown graphically in Figure 11, in which the cumulated or integral curves are plotted.

6. *Excess mortality.*—Early in March, 1919, the Census Bureau issued a supplement to its Weekly Health Index showing for 34 of

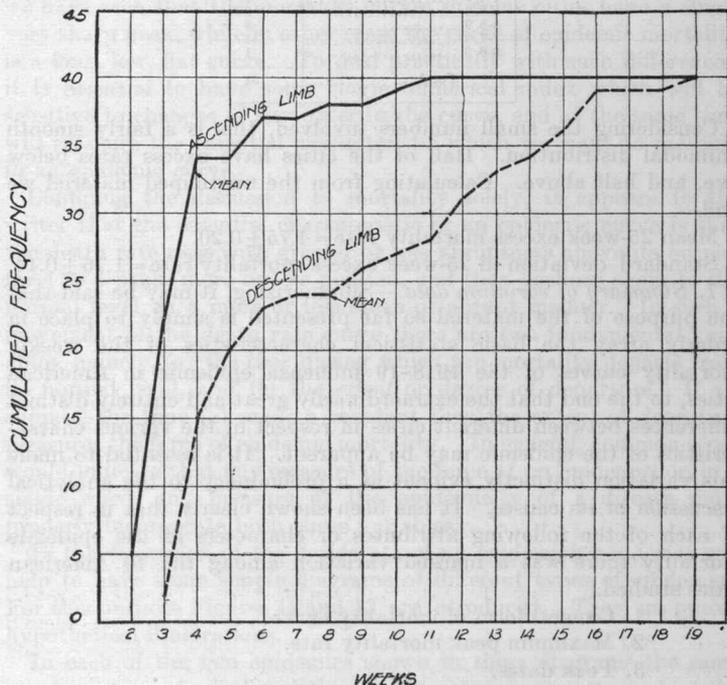


FIG. 11.—Cumulated frequency curves for time covered by (a) ascending limb, and (b) descending limb of epidemic mortality curve.

the 40 cities of Table 1 the mean excess rate of mortality due to the epidemic for the period of 25 weeks preceding March 1. These data are given in the last column of Table 1. They are arranged in the form of a frequency distribution in Table XII.

TABLE XII.—*Excess mortality for 25-week period.*

Mean excess mortality rate.	Number of cities.
1-1.9.....	1
2-2.9.....	6
3-3.9.....	6
4-4.9.....	4
5-5.9.....	9
6-6.9.....	3
7-7.9.....	4
8-8.9.....	1
Total.....	31

Considering the small numbers involved, this is a fairly smooth unimodal distribution. Half of the cities have excess rates below five, and half above. Calculating from the ungrouped material we find—

Mean 25-week excess mortality rate = 4.75 ± 0.20 .

Standard deviation in 25-week excess mortality rate = 1.76 ± 0.14 .

7. *Summary of variation data.*—Summarizing, it may be said that the purpose of the material so far presented is simply to place in orderly array the basic statistical characteristics of the weekly mortality curves of the 1918-19 influenza epidemic in American cities, to the end that the extraordinarily great and entirely distinct differences between different cities in respect of the various characteristics of the epidemic may be apparent. It is essential to make this variation distinctly evident as a preliminary to the analytical discussion of its causes. It has been shown clearly that in respect of each of the following attributes or characters of the epidemic mortality there was a marked variation among the 40 American cities studied.

1. General form of mortality curve.
2. Maximum peak mortality rate.
3. Peak dates.
4. Number of distinct peaks in mortality curve.
5. Time between peaks of mortality.
6. Steepness of ascending and descending limbs of mortality curve.
7. Excess mortality rate.
8. Duration of epidemic mortality.

The variation among cities in these different epidemiological characters constitutes a problem of first-class hygienic interest and importance. Why did it exist? Why were not all cities at least reasonably alike in their influenza epidemic? If we can find sound and correct, even though only partial, answers to these questions we shall have gained greatly in that understanding of the epidemiology of influenza which must always underlie any effective control of it. It is to the analysis of this problem that attention will next be devoted.

IV. Epidemicity Indices.

With the variation data in hand one further step is necessary before the analysis by multiple correlation can be completed. We must have a single numerical measure or index of the force of the epidemic explosion in any particular place. In the earlier sections we have seen that the mortality curves in some cities have a single very sharp peak, while in other cases the curve of epidemic mortality is a long, low, flat curve. To deal practically with such differences, it is essential to have some single numerical index which will be sensitive to changes of any order in the curve, and at the same time will measure the essential characteristic which we want to measure in an epidemic curve.

Confining the discussion to mortality solely, it appears to the writer that the essential characteristic of an epidemic curve is that the death rate rises with greater or less abruptness above its normal level to a peak, more or less pointed, and then declines again to the normal level, in a more or less steep or abrupt manner. In such a movement of the death rate curve there are two fundamental variables, namely, (a) the *time* during which the mortality departs from its normal level, and (b) the *extent* or degree of departure. If we suppose the time (a) made a constant then the extent of departure measures the force of epidemic mortality. In general, common sense would indicate that any measure of the force of an epidemic, or, in a single word, any measure of the epidemicity of a disease must properly incorporate both these variables.

In the discussion of the desiderata of an epidemicity index it will help to have some simple diagrams of different types of epidemics. For this purpose Figures 12 and 13 are introduced. They are purely hypothetical illustrations.

In each of the two epidemics shown in these diagrams the same number of people died and the peak death rate was reached at the same time. But clearly the outbreak depicted in Figure 12 would be generally regarded as a more severe or explosive epidemic, *qua* epidemic, than the one shown in Figure 13. Such changes of the death rate as are shown in Figure 13 may indeed not be regarded as epidemic at all. We do not commonly think of the seasonal rise in the endemic influenza rate as an epidemic. Yet it is quantitatively of the same order as the circumstances depicted in Figure 13. It is of the essence of the idea of an epidemic, as commonly held, that it should have something of an explosive character—that is, there must be a relatively large increase in the death (or morbidity) rate, occurring in a relatively short space of time, in order to constitute an epidemic.

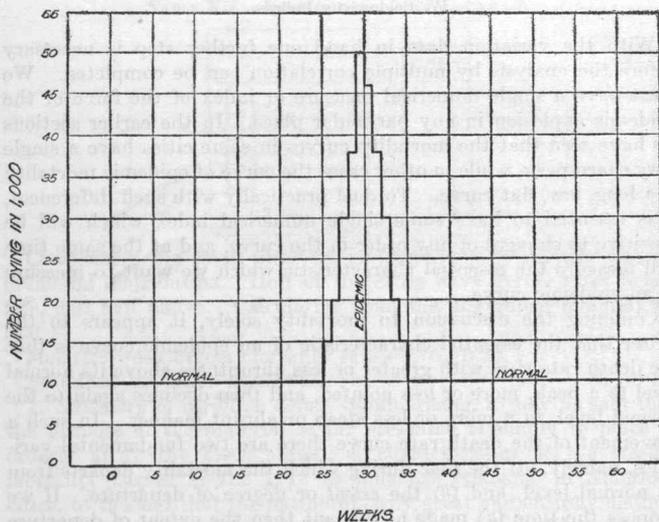


FIG. 12.—Hypothetical diagram to show epidemic of great explosiveness.

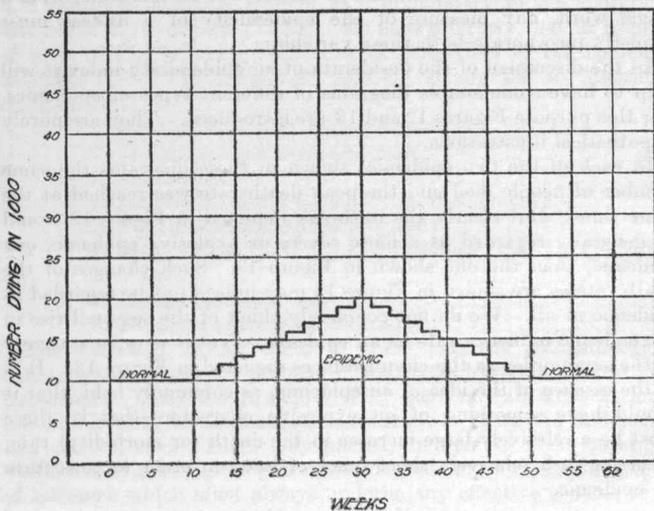


FIG. 13.—Hypothetical diagram to show epidemic of small explosiveness.

This being so, any proper measure of the degree of epidemicity must first of all measure the degree of *explosiveness* of the outbreak of the disease under discussion. There are a number of ways, mathematically, in which this can be done. The decision as to which is the best method will turn upon the degree of sensitiveness with which each measures the essentially explosive feature of the outbreak. In arriving at a measure of epidemicity for the analytical study of the influenza epidemic in American cities five different plans have been tried. We may now discuss these different indices, and decide upon which is the best for present purposes. The data used are the weekly mortality rates for thirty-nine American cities dealt with in earlier sections.

1. *Standard deviation of epidemic.*—The first epidemicity index which would occur to the biometrician is that expressed by the standard deviation of the epidemic outbreak, measured in weeks, the death rates being regarded as frequencies. An epidemic curve like that of Figure 12 obviously has a smaller standard deviation in time than one such as is shown in Figure 13. In general, the greater the explosiveness of the outbreak the smaller will be the standard deviation. Practically the manner in which this index is calculated is as follows:

(a) Take as the basis of calculation the duration of the epidemic outbreak as defined earlier.¹

(b) Within the range so defined calculate the standard deviation² in weeks in the ordinary way, the observed death rates being taken as ordinates.

In the present instance the constant takes this form: Let y denote the death rate in a particular week, and x the deviation of the week in which that rate occurred from the mean. Then, if I_1 denotes the epidemicity index, we have

$$I_1 = \sqrt{\frac{\sum x_n y x^2}{x_1 N}}$$

when N is the number of weeks in the epidemic period, and Σ denotes summation. This index is easy to calculate and has a definite physical meaning. Practically, it would probably be desirable if I_1 were to be used as an epidemicity index generally, to take some multiple of its reciprocal for tabling, since as the index now stands it becomes numerically smaller as the explosiveness of the epidemic becomes greater. The value $100/I_1$ would be satisfactory.

¹ Vide p. 20.

² The "standard deviation" is a well-known constant used in biometric work. It is the root-mean-square-deviation about the mean. For a detailed discussion of this constant see Yule's "Introduction to the Theory of Statistics" or any of the modern texts on elementary statistical methods.

2. *Variation of excess death rates.*—Another measure of epidemicity which may be considered is of a more complex character than the last. Its nature may be indicated symbolically as follows:

Let M = mean death rate during epidemic, the latter being delimited as to duration by the definition in an earlier section already referred to;

M' = mean death rate in the period from July 6, 1918, to outbreak of epidemic.

$M'' = M - M'$ = increase in mean death rate during epidemic.

$S = \sqrt{\frac{\sum_1^1 y^2}{n}}$, where y is the deviation of any particular week's death rate from M , and n is the number of weeks in the epidemic period. S is the standard deviation of the epidemic death rates, each equally weighted.

Then the second epidemicity index is

$$I_2 = \frac{100S}{M''}$$

This quantity will increase as the explosiveness of the outbreak increases. In ordinary biometric terminology it is the coefficient of variation of the weekly death rates in the epidemic period, referred to the mean excess rate as a base.

3. *Mean increase in death rate during epidemic.*—As a third epidemicity index we may take the quantity called M'' in the preceding section. We then have

$$I_3 = M''$$

4. *Twenty-five weeks excess rate.*—It has been suggested that the average excess weekly annual death rate for the 25 weeks ended March 1, 1919, might be used as a measure of the force of the epidemic. Indeed, it has been so used practically by various health officials. In the present connection we may designate this measure as I_4 .

5. *Peak-time ratio.*—An epidemicity index which immediately makes strong appeal by virtue of its simplicity is a constant for any mortality curve which may be called the peak-time ratio. The symbolical expression for it is:

$$I_5 = \frac{P - M}{T}$$

where P denotes the maximum peak mortality rate observed during the duration T of the epidemic, T being delimited by the definition stated earlier in this paper, and M is the quantity defined under the same symbol in section 2 above. This index increases as the explosiveness of the outbreak increases. In fact, it measures explosiveness in the most simple and direct way possible.

V. Numerical Values of Epidemicity Indices.

It is evident at once that these five indices have different degrees of validity and usefulness. Before attempting to discuss them in detail, however, it will be well to get the numerical values for each, in the case of each of the 39 cities under discussion. This is done in Table XIII.

TABLE XIII.—Showing values of different epidemicity indices of mortality in American cities during influenza epidemic of 1918.

Cities.	I_1 (weeks).	I_2 (per cent).	I_3 .	I_4 .	I_5 .
Albany.....	1.61	85.9	40.13	4.7	13.81
Atlanta.....	6.68	58.5	9.31	2.7	1.92
Baltimore.....	1.54	94.5	48.61	6.1	18.61
Birmingham.....	4.06	60.7	17.04	2.41
Boston.....	1.98	88.5	33.47	6.5	9.62
Buffalo.....	1.85	92.0	31.19	5.8	10.55
Cambridge.....	2.00	88.9	27.03	5.9	7.94
Chicago.....	1.98	72.4	24.04	3.8	6.61
Cincinnati.....	4.55	69.8	15.41	4.0	2.15
Cleveland.....	3.63	74.2	18.30	4.0	4.09
Columbus.....	3.55	56.4	14.94	3.2	2.74
Dayton.....	6.24	91.4	24.67	3.5	7.29
Fall River.....	1.66	80.9	38.70	5.8	11.92
Grand Rapids.....	3.41	65.7	8.10	1.5	1.08
Indianapolis.....	3.42	55.0	12.51	2.5	2.15
Louisville.....	4.11	78.4	15.45	3.6	3.07
Los Angeles.....	5.50	62.7	15.78	5.2	2.00
Lowell.....	1.70	71.5	34.60	5.1	10.58
Memphis.....	1.76	94.7	24.15	8.60
Milwaukee.....	4.48	57.4	11.57	2.9	1.53
Minneapolis.....	5.98	55.1	9.80	2.7	1.12
Nashville.....	1.58	72.6	39.39	7.8	13.83
Newark.....	5.70	99.0	15.34	5.1	2.81
New Haven.....	5.43	100.6	18.89	5.6	3.16
New Orleans.....	1.69	90.2	40.95	7.2	14.60
New York.....	2.19	71.2	23.29	4.7	5.67
Oakland.....	5.25	77.0	18.74	5.9	3.35
Omaha.....	4.17	69.6	18.47	2.91
Philadelphia.....	1.52	86.2	56.08	7.3	20.51
Pittsburgh.....	2.79	67.0	37.62	8.0	7.82
Providence.....	2.46	86.4	21.79	5.3	5.60
Richmond.....	1.33	66.1	35.12	13.91
Rochester.....	4.48	79.2	13.94	2.7	2.62
St. Louis.....	4.06	59.1	13.47	3.0	2.11
St. Paul.....	5.12	37.8	11.31	3.3	1.43
San Francisco.....	5.06	78.4	26.50	7.5	4.49
Syracuse.....	2.09	94.2	30.77	8.97
Toledo.....	1.67	69.8	17.19	2.1	5.95
Washington.....	1.49	66.3	45.08	6.6	15.34

Of these five indices there are only two which need to be taken seriously into account as practical working measures of epidemicity. These are the first and last, I_1 and I_5 . The other three fail in that they do not adequately take account of the time or duration variable, which, as we have already seen, must be an essential factor in measuring epidemic explosiveness. These other indices really measure other aspects of the epidemic better than they do explosiveness of the outbreak, which is the thing we are just now interested in. The inadequacy of I_2 , I_3 , or I_4 to measure relative explosiveness of outbreak can be readily seen by comparing, city by city, the values given in these columns of Table XIII with the curves for the same cities in Figures 1-6.

As between I_1 and I_5 the advantage, for present purposes, of I_5 is clear. It is numerically more sensitive to changes in the epidemic mortality curves. This fact is reflected in a comparison of the relative variation of the five indices which is made in Table XIV. For comparing the relative sensitivity of the indices to differences in the epidemic mortality curves, the ratio of the standard deviation of each index to its mean has been taken. This ratio has no significance in this case except for comparative purposes.

TABLE XIV.—Relative sensitivity of different epidemicity indices.

Index.	Ratio of S. D. to mean.
I_1	0.49
I_218
I_349
I_437
I_577

By conventional biometric standards it might seem *a priori* that I_1 would be a better epidemicity index than I_5 . Practically it is seen from Table XIV that the superiority of I_5 is outstanding. The reason for this superiority appears upon analysis to be that this index relates in the simplest mathematical manner possible the two essential factors in relative explosiveness, namely, the height of the explosion, and the time it required, and is therefore most sensitive to differences in relative explosiveness. The same type of constant might be used for the measure of variation in frequency curves generally, except for the fact that ordinarily it is impossible to delimit the range by absolute definition, as can be done in the case of epidemics. In an ordinary frequency curve the probable error of any determination of the range is large. The nature of the definition of the range or duration which we have here adopted for epidemic curves, as well as the characteristics of epidemic curves themselves, largely reduces this probable error in the present connection. And in any case, whatever effect the probable error of the empiric determination of duration may have will tend to be greater in the case of I_1 than of I_5 .

Taking all the facts into consideration it has been decided to adopt I_5 as the measure of explosiveness of outbreak in the further analytical study of the influenza epidemic.

VI. The Correlation of the Explosiveness of the Outbreak of Mortality in the Influenza Epidemic with Various Other Factors.

We come now to the most essential part of the study, namely, the attempt to find factors directly related to or concerned in the production of the extraordinary differences between different cities in respect of the relative explosiveness of the outbreak of epidemic mortality. The method of analysis which will be followed is that of

multiple correlation.¹ The general principle of the correlation method is simple. If in the present case, for example, we should find that, in general, when a city had a high influenza epidemicity index it also had a high density of population, and conversely, that cities having low epidemicity indices had low density of population, it would be said that there was a positive correlation in variation between explosiveness of epidemic and density of population.

In a system of n variables correlation between any two, with the others remaining constant, is measured by the coefficient

$$r_{12.34 \dots n} = \frac{r_{12.34 \dots (n-1)} - r_{1n.34 \dots (n-1)} \cdot r_{2n.34 \dots (n-1)}}{(1 - r_{1n.34 \dots (n-1)}^2)^{\frac{1}{2}} (1 - r_{2n.34 \dots (n-1)}^2)^{\frac{1}{2}}}$$

and a coefficient of zero order is found from the observations by the following well-known expression:

$$r_{12} = \frac{S(xy)}{N\sigma_1\sigma_2}$$

In the present case, because of the statistically small number of cities for which data are available, the zero order coefficients were all determined by the direct product-moment method, without the formation of correlation tables.

The first group of phenomena of which one would naturally wish to know the extent to which they were correlated with explosiveness of outbreak are certain general demographic characteristics of the several cities. The following will be considered:

(a) *Density of population.*—It is conceivable—not to say *a priori*, rather probable—that the explosiveness of outbreak of any epidemic disease would be highly correlated with the number of persons living on a unit of area. The figures for density used were calculated in terms of persons per acre of land area, on July 1, 1916.²

(b) *Geographical position.*—It is a well known epidemiological fact that, in certain classes of epidemic disease at least, the force of the epidemic diminishes as one passes from the primary center or focus. This fact was very clearly demonstrated for the 1916 poliomyelitis epidemic by Lavender, Freeman, and Frost,³ where New York City was the center. Now, in point of time, the influenza epidemic of the autumn of 1918 in the United States began in and about Boston, Mass. A great explosive outbreak occurred in Boston and Cambridge earlier than in any other cities in the country. We may then ask this question: Did the influenza epidemic, as it spread over the whole country, follow the epidemiological rule already referred to becoming less intense and less explosive the farther, geographically, it traveled from the Atlantic seaboard in general, and Boston in

¹ Public Health Bulletin No. 91, U. S. Public Health Service, 1918.

² Cf. Yule, G. U. "On the Theory of Correlation," Jour. Roy. Stat. Soc., Vol. LX, 1897, and "On the Theory of Correlation for any Number of Variables, treated by a New System of Notation," Proc. Roy. Soc. A, vol. 79, pp. 182-193, 1907.

³ Data from "Financial Statistics of Cities Having a Population of over 30,000 in 1917." Bureau of the Census, 1918.

particular? To answer this question, so far as the epidemic mortality records of the present group of cities is concerned, we have correlated the epidemicity index I_e for each city with the distance in a straight line of the same city from Boston, Mass., measuring these straight line distances on a map. Such distance measurements are rough, of course, from an absolute standpoint, but relatively they are sufficiently accurate, and may be relied on, to show correlation if any exists.

(c) *Age distribution of population.*—In the case of a disease showing so selective a mortality in respect of age as does influenza it might well be the case that the explosiveness of the outbreak of epidemic mortality would be markedly influenced by the age composition of the population in the several cities. To test this point by the correlation method one must have a single numerical measure or index of the age composition of the population in each city. Such a single numerical measure is not at hand. The problem of obtaining one is a problem which has bothered vital statisticians for a long time, as the need for it always arises in death rate correlation studies of any sort. Theoretically, of course, no *single* numerical expression can possibly be found which will uniquely describe all the properties of a complex curve. The best that can be done is some form of approximation.

For present purposes an index of differences in age composition of populations was adopted, which is admittedly rough and in special cases may be inexact, but which practically has been found, in the case of the 40 cities here dealt with, to give a sufficiently accurate picture of the differences in age constitution. The statistical procedure adopted was to determine for each city the following value:

$$\chi^2 = S \left(\frac{\Delta^2}{P} \right)$$

where Δ is the deviation for each of six age groups (viz, 0-4, 5-14, 15-24, 25-44, 45-64, 65 and over) of the percentage of the actual population of each city in 1910 in each age group, from the percentage in the same group in the Standard Population of Glover's¹ Life Table, denoted in the formula by P . S denotes summation of all six values. The value χ^2 measures through the extent to which each city deviates in the age constitution of its population from a fixed standard, but does not tell the nature or kind of the deviation. For present purposes the latter point is unessential. We are proposing to measure the correlation between explosiveness of epidemic and departure of population from normal in age distribution. Are large variations in explosiveness generally associated with large deviations in age constitution of the population? This question can be answered perfectly by the use of the present index of age consti-

¹ Glover, J. W. United States Life Tables, 1910. Bureau of the Census, 1916.

tution. If it were found that there existed a high correlation between I_5 and χ^2 it would be desirable and necessary to analyze further the nature of the deviations in age constitution. But as will presently appear this necessity does not arise.

As has been said, the age distributions for the cities in the year 1910 were used. This was necessitated by the fact that no later census data were available. It seems fairly certain, however, in as old, large, and settled communities as these dealt with are, that the age composition of the population will only change slowly, and that 1910 figures may be taken as reasonably indicative of present conditions in respect to this matter.

(d) *Percentage growth of population between 1900 and 1910.*—It might conceivably be the case that the explosiveness of the outbreak of an epidemic disease would be influenced by the rapidity with which a city had grown in the recent past. To test this possible factor in the present case the epidemicity index I_5 is correlated with the percentage growth of the population in each city in the decade 1900-1910.

The data for these various correlations are assembled in Table XV.

TABLE XV.—Data for correlation of demographic characteristics of cities with explosiveness of epidemic influenza mortality.

City.	Epi- demicy. Index I_5 .	Density of popu- lation (persons per acre).	Geo- graphical position.	Age distribu- tion χ^2 .	Growth in popu- lation.
Albany.....	13.81	8.89	128	4.76	6.5
Atlanta.....	.92	11.42	920	13.03	72.3
Baltimore.....	18.61	30.57	348	6.81	9.7
Birmingham.....	2.41	5.68	1,028	15.80	245.4
Boston.....	9.62	27.36	-----	7.18	19.6
Buffalo.....	10.55	18.97	376	8.85	20.2
Cambridge.....	7.94	28.23	3	6.51	14.1
Chicago.....	6.61	20.28	828	11.45	28.7
Cincinnati.....	2.15	9.10	712	6.73	11.6
Cleveland.....	4.09	20.08	532	11.88	46.9
Columbus.....	2.74	15.18	616	8.35	44.6
Dayton.....	7.23	12.65	684	6.56	36.6
Fall River.....	11.92	5.91	45	10.87	13.8
Grand Rapids.....	1.68	11.85	720	6.17	28.6
Indianapolis.....	2.15	10.96	776	7.23	38.1
Louisville.....	3.07	16.61	796	7.57	9.4
Los Angeles.....	2.00	2.40	2,520	7.67	211.5
Lowell.....	10.58	13.63	23	7.35	11.9
Memphis.....	8.60	12.06	1,104	14.24	28.1
Milwaukee.....	1.53	26.92	832	10.33	31.0
Minneapolis.....	1.12	11.27	1,084	11.46	48.7
Nashville.....	13.83	10.11	924	9.19	36.5
Newark.....	2.81	27.52	132	10.19	41.2
New Haven.....	3.16	13.06	100	6.81	23.7
New Orleans.....	14.60	2.96	1,332	9.25	18.1
New York.....	5.67	29.54	164	11.79	38.7
Oakland.....	3.35	6.41	2,604	6.51	124.3
Omaha.....	2.91	8.54	1,248	10.83	21.0
Philadelphia.....	20.51	21.02	260	7.19	19.7
Pittsburg.....	7.82	22.81	456	11.53	18.2
Providence.....	5.60	22.35	40	6.88	27.8
Richmond.....	13.91	10.76	460	10.55	50.1
Rochester.....	2.62	18.62	328	6.97	34.2
St. Louis.....	2.11	19.36	1,004	9.51	19.4
St. Paul.....	1.43	7.40	1,072	12.70	31.7
San Francisco.....	4.49	17.55	2,624	12.65	21.6
Syracuse.....	8.97	13.34	248	6.21	26.6
Toledo.....	5.95	10.91	620	7.26	27.8
Washington.....	15.34	9.55	376	6.58	18.8

As a matter of record, and for reference in connection with the correlation data, the mean and standard deviation of the variables included in Table XV are given in Table XVI.

TABLE XVI.—Constants for demographic data of Table XV.

Character.	Mean.	Standard deviation.
Epidemicity index, I_5	6.78 ± 0.56	5.22 ± 0.40
Density of population.....	15.17 ± .82	7.56 ± .58
Geographical position.....	721 ± 71.00	653.95 ± 50.00
Age distribution, χ^2	9.063 ± .28	2.609 ± .20
Growth in population.....	40.43 ± 5.2	48.31 ± 3.7

Coming now to the consideration of the correlations we have the following results:

(a) For the correlation between explosiveness of epidemic mortality (I_5) and density of population—

$$r = +0.092 \pm 0.107.$$

The coefficient is less than its probable error, or is, in short, substantially zero. This value justifies the conclusion that relative density of population in these 39 cities had nothing to do with the explosiveness of the influenza outbreak.

The insignificant degree of correlation in this case is shown graphically in Figure 14. The plan of this figure is first to convert the absolute values of the epidemicity index and density of population for each city to relative figures, the mean for all cities being taken as the base 100. The cities are then arranged in descending order of relative epidemicity index (solid line) and the relative density figures for the same cities are plotted as a broken line. The higher the correlation the more closely will the two lines tend to parallel each other. Here it is evident that the density line runs quite independently of the epidemicity line.

(b) For the correlation between I_5 and geographical position, measured by straight line distance from Boston

$$r = -0.348 \pm 0.095.$$

This, clearly, is a wholly different order of result from that which we had in the case of the density of population. The coefficient in the present case is nearly four times its probable error and may almost certainly be regarded as significant. The odds against its being simply a widely deviant chance result of random sampling are more than 78 to 1.¹ The sign of the coefficient is negative. This result means that the greater the linear distance of a city from Boston the

¹ Cf. Pearl, R., and Miner, J. R. A Table for Estimating the Probable Significance of Statistical Constants. *Me. Agr. Expt. Stat. Ann. Rept.* 1914, pp. 85-88.

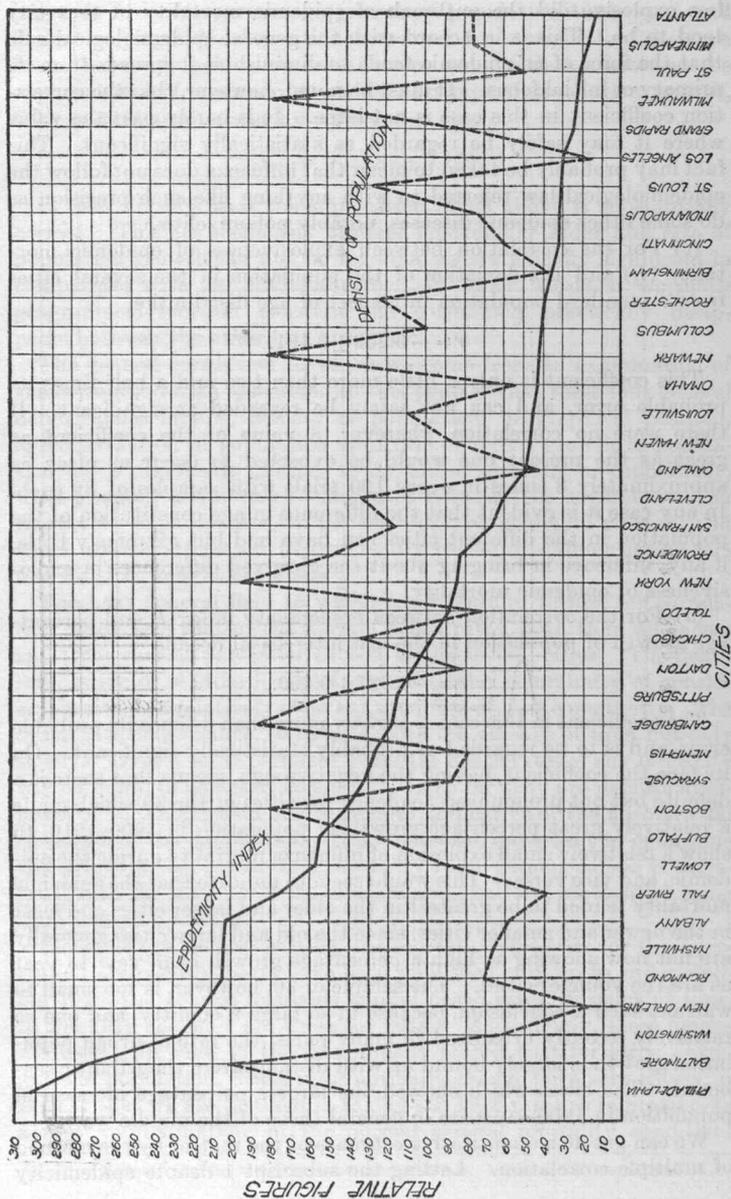


FIG. 14.—Diagram showing lack of correlation between variation in epidemicity index and density of population (Explanation in text).

less explosive did the outbreak of epidemic mortality in that city tend to be. This is in accord with the general epidemiological rule that the force of an epidemic tends to diminish as it spreads from its primary or initial focus. It must be noted, however, that the correlation coefficient in this case is not large. It is barely past the value where it may safely be regarded as statistically significant. This fact may probably be taken to mean that influenza does not follow the epidemiological law referred to with anything like such precision as do some other epidemic diseases, notably poliomyelitis.

(c) For the correlation between explosiveness of epidemic mortality (I_5) and the deviation of the population in the several cities from a standard population in respect of age distribution

$$r = -0.262 \pm 0.101.$$

This coefficient is only a little more than two and a half times its probable error, and can not safely be regarded as significant. If there were no correlation whatever, a value of the coefficient as great as the present one would be expected to occur as often as approximately 8 times in every 100 trials with samples of 39 each. In any case it is evident that the difference in age constitution of the population in the different cities can have had but extremely little, if any, influence in bringing about the observed differences in explosiveness of epidemic mortality.

(d) For the correlation between epidemicity index I_5 and percentage growth of population in the last intercensal decade

$$r = -0.327 \pm 0.096.$$

The coefficient in this case is slightly more than 3 times its probable error, and is to be regarded as probably statistically significant. On its face the coefficient, having the negative sign, means that there is a definite but not pronounced tendency for cities in the 39 which made a relatively great percentage growth in population in 1900-1910, to show a relatively small explosion of influenza mortality during the epidemic, and vice versa. This would seem to indicate that the epidemic mortality tended to be greatest in the older and larger cities and least in the newer and smaller cities, since the old and large cities generally are not now showing so high a percentage growth from year to year as are the younger cities. The sample of 39, however, is too small to warrant such a conclusion, because in so large a country, and one so relatively recently urbanized in many parts, the rate of urban population growth is largely bound up with distance from the Atlantic seaboard. The cities which showed the largest percentage increase in population in 1900-1910 are in general those of the middle west.

We can get at a quantitative estimate of the matter by the method of multiple correlation. Letting the subscript 1 denote epidemicity

index I_5 , 2 denote percentage growth of population 1900-1910, and 3 denote geographical position measured by straight line distance from Boston, as before, we have for the net correlation between the explosiveness of epidemic mortality and rate of population growth, with geographical position constant

$$r_{12.3} = -0.188 \pm 0.104.$$

It then appears that the supposition made above is substantially correct. This net coefficient between epidemicity index and rate of population growth can not be regarded as statistically significant in comparison with its probable error. In other words, if we make geographical location constant the correlation practically disappears between the other two variables.

The general conclusion to which we come from an examination of the correlation data assembled to this point is that these four general demographic factors, density of population, geographical position, age distribution of population, and rate of recent growth in population, have practically nothing to do, either severally or collectively, with bringing about those differences between the several cities in respect of explosiveness of the outbreak of epidemic mortality in which we are interested. Significantly casual or differentiating factors must be sought elsewhere.

The next general field to which one naturally turned for correlation study was that of the normal death rates, both from all causes and from various particular causes, in the several cities. The death rate, crude or standardized, of any particular community of considerable size, is a relatively constant attribute of that community. The death rate does change, to be sure, with the passage of time, but only slowly. Over a short period of years the death rates of any large city will be found to be nearly constant. In so far they are definite attributes of the city, which are, in general, indicative of the normal vital condition of the population. It is, therefore, important to determine the extent which the normal mortality from various causes is correlated with the severity of the unusual and explosive mortality arising from a great epidemic.

Since, at the time of writing, the mortality statistics for the registration area and its parts have been published only up to and including 1916, the nearest available annual death rates, in point of time, to the 1918 epidemic are those for 1916.¹ Accordingly, these figures are used. In view of the fact already stated that for large aggregates of population, death rates normally change only very slowly, it is clear that we are justified in taking the 1916 rates as indicative, to a first approximation, of the normal general mortality conditions

¹ Mortality Statistics 1916, Seventeenth Annual Report. Bureau of the Census, 1918.

prevailing in the several cities at about the time (in a broad sense) that the influenza epidemic broke out. The causes of death selected for correlation purposes in the first study are exhibited in Table XVII. For convenience of reference and comparison the epidemicity index I_5 , with which these death rates are to be correlated, is given in the second column of the table. All the death rates are crude rates.

TABLE XVII.—Data for correlation of explosiveness of influenza epidemic mortality, with death rates from various causes for 1916.

City.	Epidemicity index I_5 .	Death rate from all causes per 1,000.	Death rates per 100,000 living, from—							
			Pulmonary tuberculosis.	Organic heart diseases.	Acute nephritis and Bright's disease.	Influenza.	Pneumonia (all forms).	Typhoid fever.	Cancer.	Measles.
Albany.....	13.81	19.3	208.5	225.8	197.2	35.8	161.3	7.5	120.8	24.5
Atlanta.....	4.92	15.3	117.0	110.2	158.5	14.7	141.2	22.0	63.5	1.6
Baltimore.....	18.61	18.1	200.5	193.2	174.3	21.5	235.7	18.1	106.7	5.4
Birmingham.....	2.41	14.1	173.9	84.7	85.8	13.2	137.5	43.5	56.1
Boston.....	9.62	16.9	145.0	220.4	102.6	11.2	210.8	3.4	115.8	14.5
Buffalo.....	10.55	16.1	142.8	170.1	127.0	10.2	166.3	10.9	100.7	15.8
Cambridge.....	7.94	13.5	172.6	191.2	70.8	9.7	159.3	1.8	112.4	-7.1
Chicago.....	6.61	14.5	132.8	159.9	107.2	11.7	158.1	5.2	91.3	5.4
Cincinnati.....	2.15	16.4	208.3	202.7	158.8	26.8	145.4	3.2	116.2	15.3
Cleveland.....	2.15	14.8	132.2	119.6	90.9	16.3	132.2	7.3	83.8	8.9
Columbus.....	2.74	15.5	125.2	156.4	90.3	33.5	155.9	13.0	100.5	15.8
Dayton.....	7.20	15.2	121.8	180.8	119.5	18.9	146.2	19.7	114.8	1.6
Fall River.....	11.92	17.0	161.3	158.9	105.9	24.1	243.8	10.9	91.9	30.4
Grand Rapids.....	1.68	12.2	64.7	134.8	88.9	9.4	70.2	16.4	88.1	2.3
Indianapolis.....	2.15	15.6	159.6	175.6	115.0	17.4	141.8	26.1	99.4	9.8
Louisville.....	3.07	15.0	159.9	145.7	154.0	33.1	146.9	13.4	83.7	2.1
Los Angeles.....	2.00	12.3	176.7	161.0	111.3	9.3	78.0	2.6	105.6	2.0
Lowell.....	10.58	17.3	103.3	161.6	89.2	14.1	178.4	11.5	85.7	25.6
Memphis.....	8.60	19.8	262.1	145.1	171.1	37.0	136.9	26.7	86.2	2.7
Milwaukee.....	1.53	12.7	78.8	102.9	78.9	15.8	154.2	15.3	92.8	27.7
Minneapolis.....	1.12	12.4	117.3	120.0	101.8	8.8	111.4	5.5	96.0	20.4
Nashville.....	13.83	17.2	201.8	211.2	132.8	25.0	152.6	37.1	77.6	.9
Newark.....	2.81	15.0	145.5	153.6	140.9	17.4	161.2	6.1	85.6	25.7
New Haven.....	3.16	17.0	95.5	175.0	122.3	37.4	225.1	8.7	116.2	5.3
New Orleans.....	14.60	18.4	259.0	207.4	231.1	26.9	117.3	23.1	93.1	3.5
New York.....	5.67	13.9	154.9	168.7	131.4	9.8	179.9	3.9	84.5	9.9
Oakland.....	3.35	10.5	94.2	189.3	89.1	8.6	75.5	4.0	89.6
Omaha.....	2.91	14.4	101.5	93.7	91.3	18.7	173.4	5.0	90.0	1.8
Philadelphia.....	20.51	16.2	170.6	197.4	177.7	24.0	172.2	7.6	101.1	6.6
Pittsburgh.....	7.82	17.4	110.7	144.7	92.0	26.6	331.0	9.0	89.8	23.7
Providence.....	5.60	15.8	134.1	167.5	142.4	25.9	174.1	5.1	100.0	25.1
Richmond.....	13.91	19.7	187.0	189.5	204.9	20.4	194.0	23.6	97.0	26.2
Rochester.....	2.62	14.4	91.9	192.3	136.7	8.9	121.6	5.0	114.7	8.1
St. Louis.....	2.11	14.9	129.0	144.6	176.8	22.8	173.5	9.4	95.3	8.8
St. Paul.....	1.43	11.3	99.1	122.6	92.6	9.3	80.5	5.7	87.0	7.3
San Francisco.....	4.49	15.4	169.4	250.7	135.3	4.1	129.0	3.5	135.1	1.3
Syracuse.....	8.97	15.2	83.0	201.1	112.5	10.9	154.3	12.2	119.5	5.5
Toledo.....	5.95	18.1	168.1	192.8	89.3	19.7	156.5	22.2	97.9	33.8
Washington.....	15.34	17.8	187.4	230.5	168.1	24.2	164.3	12.9	107.7	2.2

The basic variation constants for the data of Table XVII are assembled in Table XVIII. In the last column of the table have been placed the values of the gross or zero order correlation coefficients measuring the correlation between the epidemicity index I_5 (which we have adopted as the measure of the explosiveness of the outbreak of epidemic mortality) on the one hand, and the death rates from the several causes, on the other hand.

TABLE XVIII.—Mean and standard deviation for death rates from various causes.

Cause of death.	Mean death rate.	Standard deviation in death rate.	Coefficient of correlation between epidemicity index I_e and the death rate from the specified cause.
All causes ¹	15.55±0.24	2.21±0.17	+0.661±0.061
Pulmonary tuberculosis.....	147.50±4.94	45.73±3.49	+ .525± .078
Organic heart disease.....	168.29±4.19	38.82±2.96	+ .567± .073
Acute nephritis and Bright's disease.....	127.39±4.17	38.67±2.95	+ .507± .080
Influenza.....	18.80±.96	8.86±.68	+ .287± .099
Pneumonia (all forms).....	158.40±5.18	47.99±3.66	+ .388± .092
Typhoid fever.....	12.41±1.04	9.64±.74	+ .176± .105
Cancer.....	97.07±1.62	14.99±1.14	+ .198± .104
Measles.....	11.00±1.09	10.08±.77	+ .069± .107

¹ Death rate per 1,000; in all other cases in the table the death rate is per 100,000.

The outstanding fact which strikes one at once from this table is the high order of the correlation which exists between the explosiveness of the outbreak of epidemic mortality in these communities and the normal death rate from certain causes of death in the same communities. In the first four lines of the table the correlation coefficients range from about 6 to more than 10 times the probable errors. There can be no question as to the statistical significance of coefficients of such magnitude. On the other hand, the remaining coefficients in the table are of a distinctly lower order of magnitude, ranging from smaller than the probable error up to three or four times that value. It is clear that we have here hit upon a clue as to the basis of the observed variation in cities in respect of explosiveness of epidemic influenza mortality which will repay careful examination.

The highest correlation coefficient of all is that on the first line of the table, for the correlation of epidemicity index with death rate from all causes. The existence of this high correlation at once indicates that an essential factor in determining the degree of explosiveness of the outbreak of epidemic influenza in a particular city was the normal mortality conditions prevailing in that city. In the group of communities here dealt with those cities which had a relatively high normal death rate had also a relatively severe and explosive mortality from the influenza epidemic. Similarly, cities which normally have a low death rate had a relatively low, and not sharply explosive, increase in mortality during the epidemic.

It will also be noted that the correlations in the next three lines of the table, namely those for pulmonary tuberculosis, so-called organic diseases of the heart, and chronic nephritis and Bright's disease, are of the same order of magnitude as that between the death rate from all causes and the explosiveness of epidemic outbreak of influenza. These facts have certain aspects of general biological, and, in the

opinion of the writer, hygienic interest. They will, however, not be discussed here, save in one respect.

Because of the potential importance of these facts, it is desirable to examine them with the greatest critical care. A point which occurs to one at once is the possibility that the observed high correlation between epidemicity index and pulmonary tuberculosis, organic heart diseases, and acute nephritis and Bright's disease, arises because of differences in age constitution of the population in the different cities. In general, it is known that the crude death rate from these causes is influenced, in greater or less degree, by the age constitution of the population. May this not be the whole, or at least the main, cause of the observed correlation? Again, it has already been seen earlier in the paper that there is a distinct, though small, correlation between the geographical position of the cities studied and the explosiveness of the epidemic mortality. May this factor not play an important part in the observed correlations of the epidemicity index with the causes of death showing a high correlation with epidemicity index?

The simplest and most direct method of settling these questions is that of multiple correlation. What is needed is to get the net correlation between the death rate from organic heart diseases, let us say, and epidemicity index, for a constant age distribution of the population and constant geographical position. In the usual terminology of vital statistics we must correct our results for age distribution and geographical position. If we let the subscript 1 denote the cause of death (pulmonary tuberculosis, organic heart disease, or acute nephritis and Bright's disease, as the case may be); the subscript 2 denote the value of the measure of the explosiveness of the epidemic mortality, our epidemicity index I_e ; the subscript 3 denote geographical position, measured as before by linear distance from Boston; and the subscript 4 denote deviation of the population from a standard age distribution, the thing desired to settle the points raised above is the net correlation coefficient, $r_{12.34}$.

By means of the equation already given (p 33) these net coefficients have been determined with the following results:

1. Net correlation between influenza epidemicity index and death rate from pulmonary tuberculosis, for constant age distribution and geographical position, $r_{12.34} = +0.609 \pm 0.068$
2. Net correlation between influenza epidemicity index and death rate from organic diseases of the heart, for constant age distribution and geographical position, $r_{12.34} = +0.594 \pm 0.070$
3. Net correlation between influenza epidemicity index and death rate from acute nephritis and Bright's disease, for constant age distribution and geographical position, $r_{12.34} = +0.510 \pm 0.080$

From these results it is seen that, instead of the correlation between the explosiveness of epidemic mortality and death rate from the

diseases mentioned being due to uncorrected age and locality factors, the *net* correlations after correction has been made for these factors, are *actually higher than were the gross, uncorrected correlations*. The net correlation of the pulmonary tuberculosis death rate with epidemicity index is the highest of the three. It has a value about 9 times its probable error. The chances are literally billions to 1 against this correlation being due to accident or chance. We may conclude that the most significant factor yet discovered in causing the observed wide variation amongst these 39 American cities in respect of the explosiveness of the outbreak of epidemic influenza mortality in the autumn of 1918 was the relative normal liability of the inhabitants of the several cities to die of one or another of the three great causes of death which primarily result from a functional breakdown of one of the three fundamental organ systems of the animal body, the lungs, the heart, and the kidneys.

VII. Summary.

In this first study the weekly mortality statistics of the influenza epidemic beginning in the autumn of 1918 have been analyzed in a preliminary way for some 39 large American cities. It has been shown in the first instance that there was an extraordinary degree of variation amongst the several cities in this group of cities in respect of the relative degree of explosiveness of the outbreak of epidemic mortality. The first problem confronting the student of the epidemic was the analysis of this variation, to find, if possible, primary factors concerned in its causation. Such an analysis, by the method of multiple correlation, appears to demonstrate that an important factor so far found in causing the observed wide variation amongst these 39 American cities in respect of the explosiveness of the outbreak of epidemic influenza mortality in the autumn of 1918 was the magnitude of the normal death rates observed in the same communities, particularly those death rates from pulmonary tuberculosis, diseases of the heart and of the kidneys.



The first part of the paper is devoted to a description of the experimental apparatus and the method of observation. The second part is devoted to a description of the results obtained in the experiments. The third part is devoted to a discussion of the results and a comparison with the results obtained by other investigators. The fourth part is devoted to a summary of the results and a conclusion.

The results of the experiments show that the rate of reaction is proportional to the concentration of the reactants. This is in agreement with the law of mass action. The rate of reaction is also proportional to the temperature. This is in agreement with the Arrhenius equation. The results of the experiments also show that the reaction is reversible. This is in agreement with the law of chemical equilibrium. The results of the experiments also show that the reaction is first order with respect to the concentration of the reactants. This is in agreement with the law of mass action. The results of the experiments also show that the reaction is first order with respect to the temperature. This is in agreement with the Arrhenius equation. The results of the experiments also show that the reaction is reversible. This is in agreement with the law of chemical equilibrium. The results of the experiments also show that the reaction is first order with respect to the concentration of the reactants. This is in agreement with the law of mass action. The results of the experiments also show that the reaction is first order with respect to the temperature. This is in agreement with the Arrhenius equation. The results of the experiments also show that the reaction is reversible. This is in agreement with the law of chemical equilibrium.

The results of the experiments also show that the reaction is first order with respect to the concentration of the reactants. This is in agreement with the law of mass action. The results of the experiments also show that the reaction is first order with respect to the temperature. This is in agreement with the Arrhenius equation. The results of the experiments also show that the reaction is reversible. This is in agreement with the law of chemical equilibrium. The results of the experiments also show that the reaction is first order with respect to the concentration of the reactants. This is in agreement with the law of mass action. The results of the experiments also show that the reaction is first order with respect to the temperature. This is in agreement with the Arrhenius equation. The results of the experiments also show that the reaction is reversible. This is in agreement with the law of chemical equilibrium.