Methods for Current Statistical Analysis of Excess Pneumonia-Influenza Deaths

ROBERT E. SERFLING, Ph.D.

THE USE of excess mortality as an index for recognition of influenza outbreaks is well known through the detailed studies of the late Dr. Selwyn Collins. In his historical reviews, the sequence of influenza outbreaks in the United States was traced back to the late 19th century, and in analytical studies age differentials and other factors were investigated. A bibliography of Collins' many papers in this field was recently published (*1*).

Collins' primary, if not sole, interest in excess mortality was with its use as an analytical tool for the definition of epidemiologic characteristics of influenza epidemics. This paper has a different objective—the use of weekly reports of pneumonia-influenza deaths by cities as an early quantitative measure of the severity of an influenza epidemic and its geographic localization. Attention is centered on statistical methods for constructing standard curves of expected seasonal mortality against which reported deaths may be compared as they occur. Retrospective analysis is mentioned only incidentally.

The emphasis given to analysis of excess pneumonia-influenza mortality in the statistical study of this disease is a result of the many problems associated with other measures of prevalence of influenza. Reporting of cases of influenza through routine channels is unsatisfactory because of the difficulty in clinical differentiation of mild influenza from other upper respiratory illness. Laboratory diagnostic tests are useful in establishing the presence of influenza but give only qualitative measures of prevalence. Epidemic reports also suffer from the absence of any numerical standard through which the severity of a particular epidemic can be determined.

Industrial and school absenteeism data provide useful information during epidemic periods but are not ordinarily available over a broad geographic network and therefore cannot be used for calculation of expected seasonal incidence in national pneumonia-influenza surveillance. In the 1957–58 influenza epidemic, the U.S. National Health Survey provided weekly estimates of prevalence of upper respiratory illness, but this information is not published during endemic periods.

Although excess pneumonia-influenza mortality lags about 4 weeks behind increases in influenza morbidity, it serves as the primary quantitative surveillance index. For this purpose, methods have been developed which differ from those of Collins in three respects:

1. Advance estimation of expected normal mortality rather than postepidemic analysis.

2. Graphic presentation of "expected seasonal mortality" and observed deaths rather than deviations from expectancy.

3. Use of total number of deaths instead of rates.

These distinctions are illustrated in figure 1, which presents a 35-year analysis by Collins, and in figure 2, a Communicable Disease Center influenza surveillance chart.

Total pneumonia-influenza deaths rather

Dr. Serfling is chief of the Statistics Section, Epidemiology Branch, Communicable Disease Center, Public Health Service, Atlanta, Ga.

than rates are used in the surveillance procedure since the data are weekly deaths by occurrence in U.S. cities. Current definition of the population in which these deaths occur is generally not possible since hospital deaths of persons from the extensive suburban fringes of rapidly growing metropolitan areas are included.

Methods

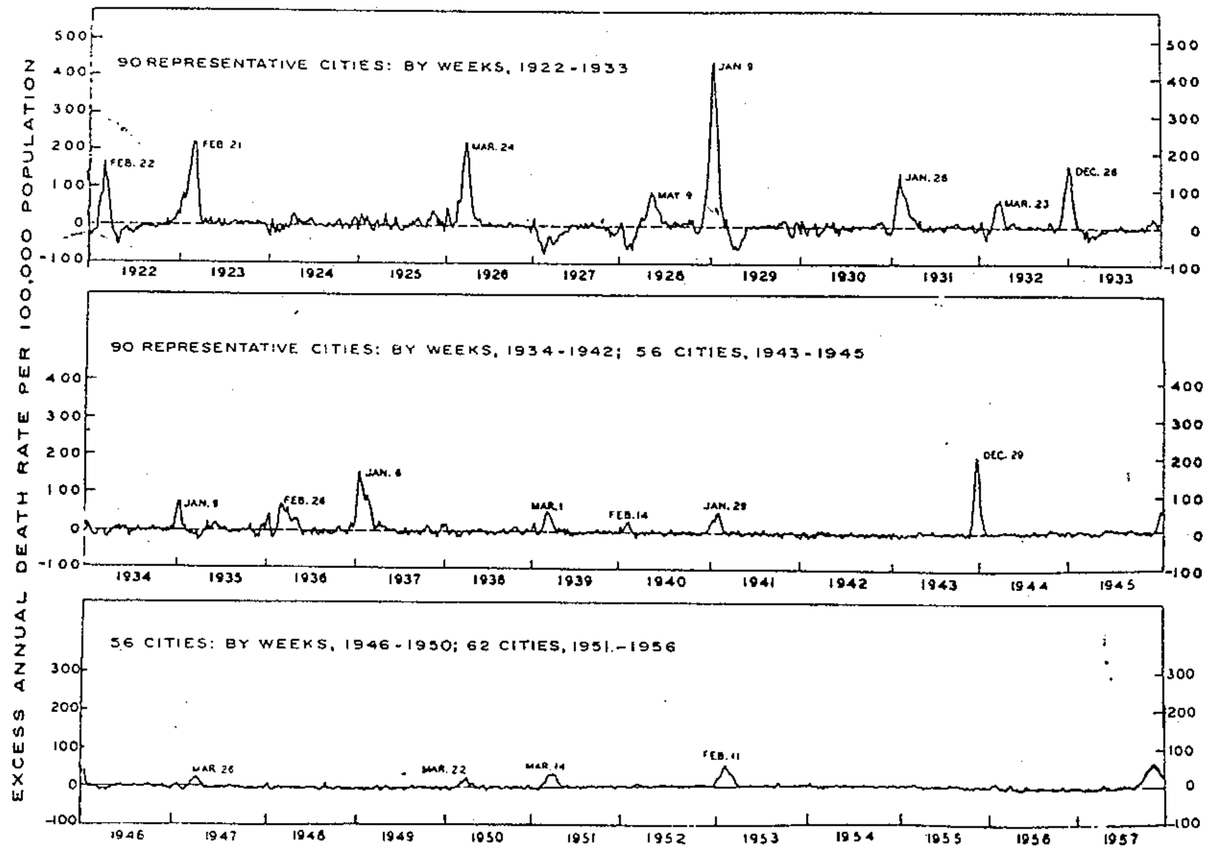
Advance estimation of expected pneumonia-influenza mortality levels requires (*a*) determination of secular trend and its extrapolation, (*b*) estimation of seasonal variation, and (*c*) distinction between epidemiologically significant departures from expected weekly mortality and random variation from endemic levels.

The first and second are essentially statistical problems for which techniques will be described. The third problem is also partly one of statistical method, but it requires a greater degree of epidemiologic insight and general knowledge of the reporting system and its idiosyncrasies.

Statistical methods used in Communicable Disease Center estimates of expected weekly mortality have been of two types: point-to-point linear estimates and Fourier series with linear trend.

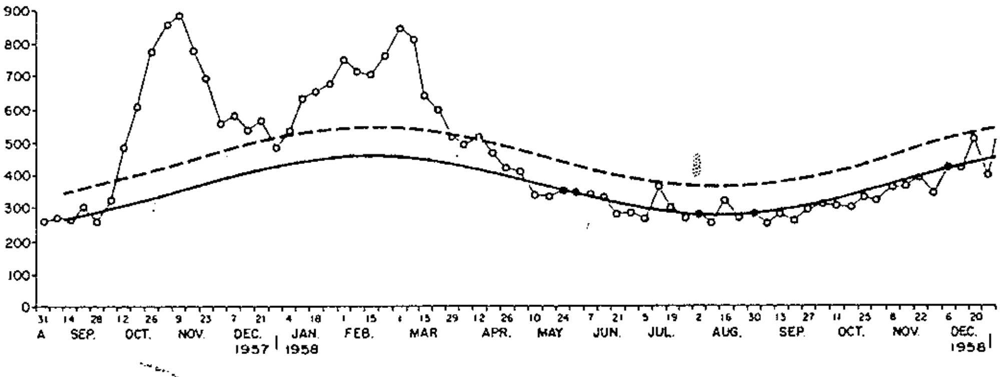
The first method was developed after a 1957 study of the secular trend of pneumonia-influenza deaths indicated that a simple linear projection of the 1954–56 levels, week by week, would provide a reasonably accurate estimate of expected 1957–58 deaths. A straight line was fitted by least squares to, for example, week 40 of calendar years 1954, 1955, and 1956, and from this line a projection was made to week 40 of 1957. The 52 weekly lines were computed with a common slope but different means after a preliminary analysis of variance showed no significant difference in slope of the 52 regres-

Figure 1. Excess annual death rate in representative U.S. cities, 1922–57



SOURCE: S. D. Collins, Public Health Monograph No. 48, 1957.

Figure 2. Weekly pneumonia-influenza deaths



sion lines. The resulting estimates were grouped into thirteen 4-week periods and the average values for each period were then connected by straight lines. Seasonal curves constructed in this manner appeared in the 1957–58 influenza surveillance reports of the Communicable Disease Center. Recently this procedure was used to obtain estimates of excess mortality during the 1957–58 and 1959–60 influenza epidemics using monthly data for the entire United States. Results are given in table 1. In this table, for example, linear estimates of expected cardiovascular-renal deaths in February 1958 and 1960 were made from a straight line fitted to the data for February of 1954, 1955, 1956, 1957, and 1959.

The point-to-point estimation procedure is simple and rapid, but yields seasonal curves which reflect distracting irregularities of the data. The moving average methods used by Collins also have this disadvantage when expected deaths are shown as a seasonal distribution curve.

Such irregularities may be removed from the curve of expected mortality by use of a mathematical function to describe seasonal change. An appropriate function was found by combination of a linear term describing secular trend with sine and cosine terms describing seasonal change to form an equation of the type

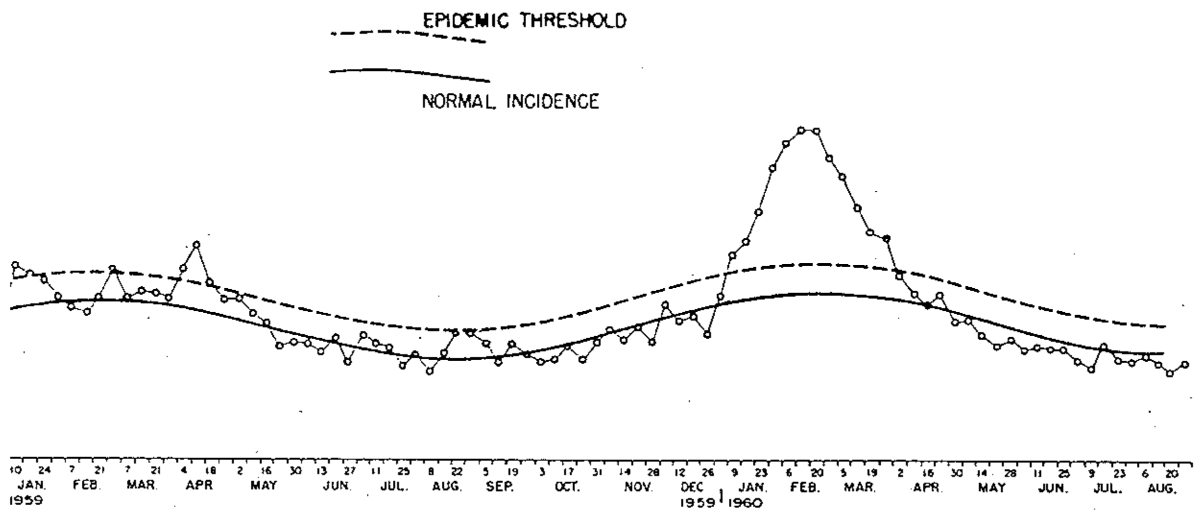
$$\hat{Y}=u+bt+\Sigma a_t \cos \theta+\Sigma b_t \sin \theta \qquad [1]$$

in which θ is a linear function of t.

A preliminary study using Bliss' analysis (*2*) of variance methods to determine the appropriate number of Fourier terms indicated that a single Fourier term might be sufficient. Further tests were made annually by computing models with two Fourier terms and with one. The following analysis of variance, using data for the 108 U.S. cities presented in figure 2, agreed with earlier studies in indicating that one Fourier term gave a satisfactory representation of seasonal change in pneumonia-influenza data for the 108 cities. Numerical values of the parameters of the one-term model are given in equation 2.

Model	*Source*	*Sum of squares*	*Degrees of freedom*	*Mean squares*
Two Fourier terms	Regression	4, 085, 701	6	---------
One Fourier term	Regression	4, 069, 429	4	---------
	Difference	16, 272	2	8, 136
Residual of two-term model		434, 189	33	13, 157

in 108 cities of the United States



Least squares estimation of the parameters in equation 1 leads to

$$\hat{Y}=300.5+2.1t+97.6 \cos\left(\frac{2\pi t}{13}-2.67\right) \quad [2]$$

for the seasonal curve for 108 U.S. cities. In this equation t is measured in units of 4-week periods with origin, $t=0$, at the midpoint of the 4-week period ending September 24, 1955.

Table 1. Estimated excess mortality by age and selected causes, United States, 1957, 1958, and 1960

Age group (years) and causes	1957 October–December	1958 January–March	1960 January–March
All ages	39, 300	20, 000	26, 700
Under 1	600	200	[1] −3, 200
1–14	1, 300	100	100
15–24	1, 000	300	200
25–44	3, 600	700	2, 200
45–64	10, 400	5, 300	6, 100
65 and over	22, 400	13, 400	21, 300
Selected causes	39, 300	20, 000	26, 700
Pneumonia-influenza	12, 100	6, 000	10, 600
Cardiovascular-renal	18, 700	13, 000	12, 200
All other causes	8, 500	1, 000	3, 900

[1] The infant mortality rate in January 1960 was the lowest rate ever recorded for January (reported in *Monthly Vital Statistics Reports*, vol. 9, No. 3, May 18, 1960, issued by National Vital Statistics Division, Public Health Service).

Data used were for the years 1955–56, 1956–57, and 1958–59, omitting the epidemic year of 1957–58.

The epidemic threshold line in figure 2 is placed at a distance of 1.64 standard deviations above the trend line, a level which experience has shown to be useful for distinguishing epidemic increase from random variation. The standard deviation at each point is corrected for linear extrapolation, but not for the varying contribution of the Fourier term. Experience with this index has shown that an elevation above the epidemic threshold for 2 or more weeks usually indicates a rise in mortality of epidemiologic interest. A high level for a single week is often accounted for by delayed reports, error in transmission of the data, or some other extraneous factor.

In equation 2 the parameters were estimated by least squares. For this model the inverse matrix can readily be obtained by algebraic methods after simplification of the coefficients of the normal equations by finite integrations. Computation by this method is not excessive, but is increased in some cases by the need for adjustment to obtain suitable values for the residual sum of squares used in estimating the standard deviation of expected levels. Adjustment becomes necessary because of the occurrence of high values for single weeks. Such artifacts greatly inflate the residual sum of squares. Adjustment is made by replacing such values with estimated trend-line values, recalculating the residual sum of squares, and then computing the standard deviation with appropriate reduction in degrees of freedom.

A disadvantage of the procedure arises from the fact that secular trend is estimated only from endemic years since epidemic years are excluded entirely. A case in point is estimation of expected seasonal incidence for 1960–61. The years 1957–58 and 1959–60 both had large influenza epidemics so that use of the model described, omitting the epidemic years, would require a 2-year projection from September 1958. In order to reduce extrapolation, a more flexible procedure enabling secular trend estimates to be made only from periods of low seasonal incidence has been developed. Four steps are required:

1. Estimation of secular trend.
2. Removal of secular trend from the data.
3. Estimation of seasonal change from the adjusted data.
4. Restoration of the trend component.

In this procedure secular trend is estimated from selected periods during the season of low incidence, approximately mid-May through mid-September. In the "pneumonia-influenza year" of thirteen 4-week periods (beginning with the 35th week of the calendar year) used throughout this paper, the low season includes the 10th period of one year through the 1st period of the next.

Within the five-period low season, some screening may be necessary in selection of periods to be used for trend estimation. Summer heat waves, for example, sometimes cause a rise in urban deaths which is reflected in the reported pneumonia-influenza deaths. If such an artifact occurs near either end of the sequence of years used for the trend estimate, it may have an appreciable effect on the slope.

The procedure for this and successive steps of the estimation procedure will be illustrated by calculation of 1960–61 estimates with use of pneumonia-influenza data for the 108 U.S. cities presented in table 2. The average number of deaths in the 10th and 11th periods of each of the 6 years 1955–60 were used to estimate the trend line, for which the least squares equation is:

$$\hat{Y}=313.83+1.549(x-43) \qquad [3]$$

Secular trend was removed from the data by taking the trend component at the center of the trend line in period 43 as zero and cumulatively adding 1.5 (the slope of the trend line) to each period preceding period 43. Thus, 1.5 was added to the number of deaths in period 42, 3.0 to the number of deaths in period 41, and so on. In this manner secular trend was

Table 2. Pneumonia-influenza deaths in 108 U.S. cities

Period	Calendar week numbers	4-week averages						
		1954–55	1955–56	1956–57	1957–58	1958–59	1959–60	1960–61
1	[1] 35–38	196	214	233	276	272	341	324
2	39–42	245	243	284	420	308	323	348
3	43–46	260	308	294	826	350	391	392
4	47–50	322	338	357	592	400	433	404
5	51–2	378	434	390	553	[2] 518	546	522
6	3–6	392	404	448	697	488	900	527
7	7–10	415	388	422	782	496	885	500
8	11–14	354	404	396	564	562	610	483
9	15–18	300	368	349	455	495	449	420
10	19–22	258	308	290	345	374	358	352
11	23–26	237	291	310	313	349	333	327
12	27–30	288	222	290	303	333	309	319
13	31–34	264	244	256	272	350	292	320
Total		3,909	4,160	4,319	6,398	5,295	6,170	5,248

[1] End of 35th week ranges from August 30 through September 5.
[2] 5-week average since 1958 includes 53 weeks when successive years are standardized to years of 52 weeks, numbered 1 through 52.

Table 3. Pneumonia-influenza deaths in 108 U.S. cities with secular trend removed, standard seasonal curve, and standard seasonal curves with secular trend restored

4-week periods	Seasonal curves with secular trend removed			Average of columns 2, 3, 4	Standard seasonal curve	Standard seasonal curve with addition of secular trend						
	1954–55	1955–56	1956–57			1954–55	1955–56	1956–57	1957–58	1958–59	1959–60	1960–61
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)
1	259.0	257.5	257.0	258	270.7	208	227	247	266	286	305	325
2	306.5	285.0	306.5	299	298.7	237	257	276	296	315	335	354
3	320.0	348.5	315.0	328	339.1	279	299	318	338	357	377	396
4	380.5	377.0	376.5	378	385.2	327	346	366	[1] 385	405	424	444
5	435.0	471.5	408.0	438	425.3	368	388	407	427	446	466	485
6	447.5	440.0	464.5	451	445.4	390	409	429	448	468	487	507
7	469.0	422.5	437.0	443	442.5	388	408	428	447	466	486	506
8	406.5	437.0	409.5	418	418.3	366	385	405	424	444	463	483
9	351.0	399.5	361.0	370	378.8	328	347	367	386	406	425	445
10	307.5	338.0	300.5	315	335.2	286	305	325	344	364	383	403
11	285.0	319.5	319.0	308	300.3	252	272	291	311	330	350	369
12	334.5	249.0	297.5	294	275.7	229	249	268	288	307	327	346
13	309.0	269.5	262.0	280	264.8	220	239	259	278	298	317	337
Total	4,611.0	4,614.5	4,514.0	4,580	4,580.0	3,878	4,131	4,386	4,638	4,892	5,145	5,400

[1] 43d 4-week period—midpoint of trend line.

removed from all periods in 1954–55, 1955–56, and 1956–57. The adjusted data were then used to estimate the curve of seasonal variation. The data with trend removed are shown in columns 2, 3, and 4 of table 3. The curve of seasonal variation can be estimated from the trend-free data by a modification of the least squares procedure for fitting a Fourier function described above.

An even simpler estimation procedure is available in the "double integration" method recently described by W. B. Langbein (*3*) as a modification of a procedure (*4*) which had been developed for a different purpose—determination of the period of long-term cyclic data rather than the description of a known cyclic period. In passing it might be noted that this method may produce spurious cycles if applied to noncyclic data. This problem does not arise in the present application since the data are clearly cyclic.

The rationale of Langbein's method is based on setting the deviations from the mean of a set of observed values equal to an appropriate sine function and integrating twice. In theory, let

$$\hat{Y}-\overline{Y}=A\sin\left[\frac{2\pi(h-1)}{N}+\Phi\right]$$

in which A is the modulus of amplitude, N the number of periods in a cycle (13 in the present example), and Φ the phase angle with $h=1, 2 \ldots N$.

Then

$$A\int_{\frac{1}{2}}^{h+\frac{1}{2}}\sin\left[\frac{2\pi(h-1)}{N}+\Phi\right]dh$$

$$=-\frac{AN}{2\pi}\cos\left[\frac{2\pi(h-\frac{1}{2})}{N}+\Phi\right] \qquad [5]$$

after subtraction of the mean value,

$$\frac{AN}{2\pi}\cos\left(\Phi-\frac{\pi}{N}\right)$$

Integrating the right hand side of equation 5 over the range $h=\frac{1}{2}$ to $h+\frac{1}{2}$ gives

$$-\frac{AN^2}{4\pi^2}\sin\left(\frac{2\pi h}{N}+\Phi\right) \qquad [6]$$

after subtraction of the mean value,

$$\frac{AN^2}{4\pi^2}\sin\Phi$$

Equation 6 is a linear function of the right-hand side of equation 4 but with h, the period number in the cycle, advanced one unit.

In practice the estimation procedure depends upon carrying out an equivalent set of numerical operations on the left-hand side of the equation

$$Y_h - \bar{Y} = A \sin\left(\frac{2\pi(h-1)}{N} + \Phi\right)$$

but with integration replaced by summation.

The procedure is rather tedious owing to the repeated subtraction of means and cumulation of positive and negative deviations which must be carried to several decimal places to reduce rounding errors. To avoid these annoyances the procedure was carried through algebraically and is presented as a computing form in table 4. A derivation of the computing formulas is given in the technical note at the end of the paper.

In table 4 the data, Y_h, are entered from column 5 of table 3, and then cumulated twice as shown in the two columns to the right of the Y_h column. The totals T_1, T_2, and T_3 of the three columns are then substituted in the computing formulas S_h and C. To illustrate the computation, the second value in the S_h+C column (for weeks 39–42) is equal to

$$26(815) - 2(64{,}782) + 22(4{,}580) + 9{,}082 = 1{,}468.$$

The corresponding entry in the next column $\hat{Y}_{h+1}$ is equal to

$$352.31 - 0.008985(1{,}468) = 339.1.$$

Each of the above operations can be carried out as a continuous operation on a desk calculator. Since no division occurs until the last operation, any desired degree of accuracy can be assured. Note that in the column $\hat{Y}_{h+1}$ each entry is advanced one period in the cycle from the initial column, Y_h.

The standard seasonal curve calculated in table 4 is shown in table 3, column 6. In columns 7–13 of table 3, secular trend has been added to the standard seasonal curve. Restoration of secular trend is accomplished by reversing the process through which it was removed.

Table 4. Computing procedure for fitting sine curve by double integration method

Week numbers	Y_h	Cumulative of Y_h		h	$h(13-h)$	S_h+C	$\hat{Y}_{h+1}$	$h+1$
		1st	2d H					
35–38	258	258	258	1	12	5, 968	298. 7	2
39–42	299	557	815	2	22	1, 468	339. 1	3
43–46	328	885	1, 700	3	30	−3, 664	385. 2	4
47–50	378	1, 263	2, 963	4	36	−8, 128	425. 3	5
51–2	438	1, 701	4, 664	5	40	−10, 364	445. 4	6
3–6	451	2, 152	6, 816	6	42	−10, 034	442. 5	7
7–10	443	2, 595	9, 411	7	42	−7, 346	418. 3	8
11–14	418	3, 013	12, 424	8	40	−2, 950	378. 8	9
15–18	370	3, 383	15, 807	9	36	1, 906	335. 2	10
19–22	315	3, 698	19, 505	10	30	5, 792	300. 3	11
23–26	308	4, 006	23, 511	11	22	8, 526	275. 7	12
27–30	294	4, 300	27, 811	12	12	9, 744	264. 8	13
31–34	280	4, 580	32, 391	13	0	9, 082	270. 7	1
Total	4, 580	32, 391	158, 076			0	4, 580. 0	
Symbols	T_1	T_2	T_3					

Y_h = average number of pneumonia-influenza deaths, 1954–55 through 1956–57, in 108 U.S. cities during 4-week period h after removal of secular trend.

General formulas:

$$\hat{Y}_{h+1} = \frac{T_1}{N} - \frac{2\pi^2}{N^3}(S_h + C)$$

$$S_h = 2NH - h(2T_2) + h(N-h)T_1$$

$$C = -T_1\left(\frac{N^2-1}{6}\right) + (N+1)T_2 - 2T_3$$

Example: $N = 13$

$$C = 9{,}082$$

$$\frac{T_1}{N} = 352.31$$

$$\frac{2\pi^2}{N^3} = 0.008985$$

$$S_h + C = 26H - h(64{,}782) + h(13-h)4{,}580 + 9{,}082$$

Table 5. Calculation of weekly expected deaths for 1960–61

Period	$\hat{Y}_i$	y_{i1}	y_{i2}	y_{i3}	y_{i4}
1955–60					
13	317				
1960–61					
1	325	319	323	328	334
2	354	341	350	359	369
3	396	379	[1] 390	402	414
4	444	426	438	450	461
5	485	472	481	489	496
6	507	501	506	508	509
7	506	509	507	504	500
8	483	494	487	479	470
9	445	460	450	440	429
10	403	418	408	398	389
11	369	381	373	365	359
12	346	353	348	344	341
13	337	339	337	337	338
1961–62					
1	345				
2	374				

[1] Example: y_{32}. $390 = 354(0.0498) + 396(1.0459) + 444(-0.1162) + 485(0.0205)$.

It will be recalled that the origin for calculation of secular trend was the 43d period of the data, which is the 4th period of 1957–58 in column 10 of table 3. The trend component is zero in this period and it will be observed that the entry in this cell, 385, is equal to the value for the fourth period of the standard curve in column 6 of table 3.

The other values of the standard seasonal curves with trend restored, columns 7–13 of table 3, were obtained by cumulative subtraction or addition of the linear trend component, 1.5 deaths per period, to the standard seasonal curve. Thus the entry 338 for the third period of 1957–58 equals 339.1 − 1.5; the entry 296 for the second period equals 298.7 − 2(1.5), and so forth. Moving forward from the fourth period of 1957–58, the values 427, 448, and 447 were obtained by adding, respectively, 1.5, 3.0, and 4.5 to the values 425.3, 445.4, and 442.5 of the standard curve in column 6.

The resulting curve of expected values is shown in figure 3. This figure also illustrates the linear secular trend, exhibits the standard seasonal curve as an insert, and shows the observed deaths in each 4-week period. Peak periods of recent epidemics are labeled with dates of the 4-week periods.

Graphic presentation of the expected and observed mortality as shown in figure 2 is sufficient for current surveillance of pneumonia-influenza mortality. For more detailed study it is useful to have weekly numerical estimates of the expected number of deaths.

Such estimates can be obtained by use of a cubic interpolation formula in which the coefficients are obtained by passing a cubic through four successive equidistant points and substituting values of the independent variable which give weekly estimates at the proper location. The coefficients used in obtaining the weekly values for 1960–61 shown in table 5 are as follows:

$\hat{Y}_i$ = Average expected number of deaths in a 4-week period i. Geometrically $\hat{Y}_i$ is located in the middle of period i.

$\hat{y}_{ij}$ = Average expected number of deaths in week j of period i. If the interval from $\hat{Y}_i$ to $\hat{Y}_{i+1}$ is unity the four $\hat{y}_{ij}$ of period i will be located at distances of $-\frac{3}{8}$, $-\frac{1}{8}$, $\frac{1}{8}$, $\frac{3}{8}$ from $\hat{Y}_i$.

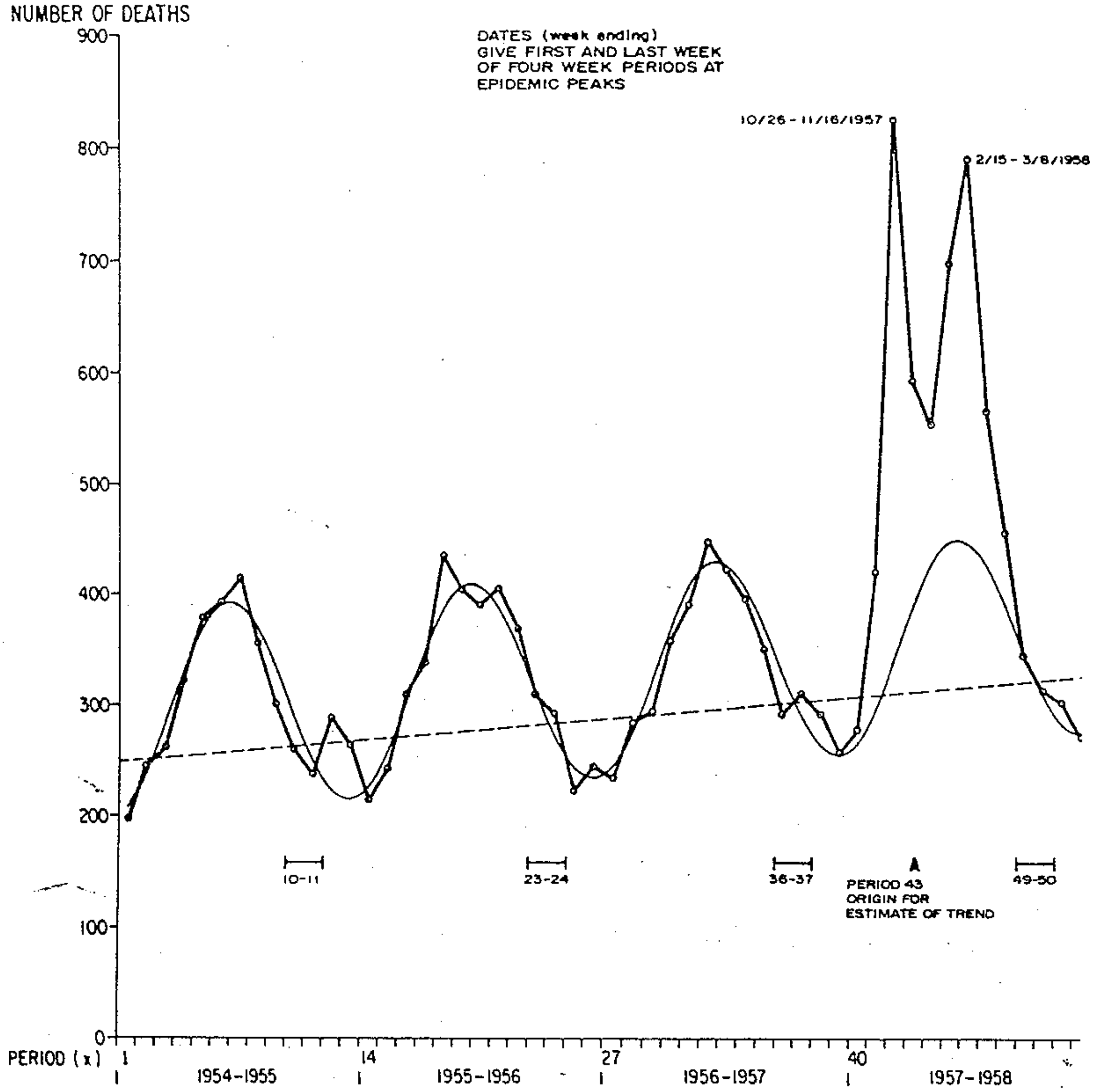
With these conventions, the following table of coefficients, k_{ij} enables calculation of the $\hat{y}_{ij}$ (expected weekly values of the ith period) from values of $\hat{Y}_i$ for four successive periods.

$$\hat{y}_{ij} = \sum_{1}^{4} \hat{Y}_i k_{ij}$$

4-week periods	k_{i1}	k_{i2}	k_{i3}	k_{i4}
$i-1$	0.2041	0.0498	−0.0342	−0.0635
i	1.0205	1.0459	.9229	.6982
$i+1$	−.2783	−.1162	.1318	.4189
$i+2$	.0537	.0205	−.0205	−.0537

To determine the epidemic threshold, an estimated variance is obtained from the sum of squares of differences between observed and expected weekly values for a nonepidemic year. Degrees of freedom will be equal to the number of observations, reduced by four in order to

Figure 3. Procedure for estimating expected number

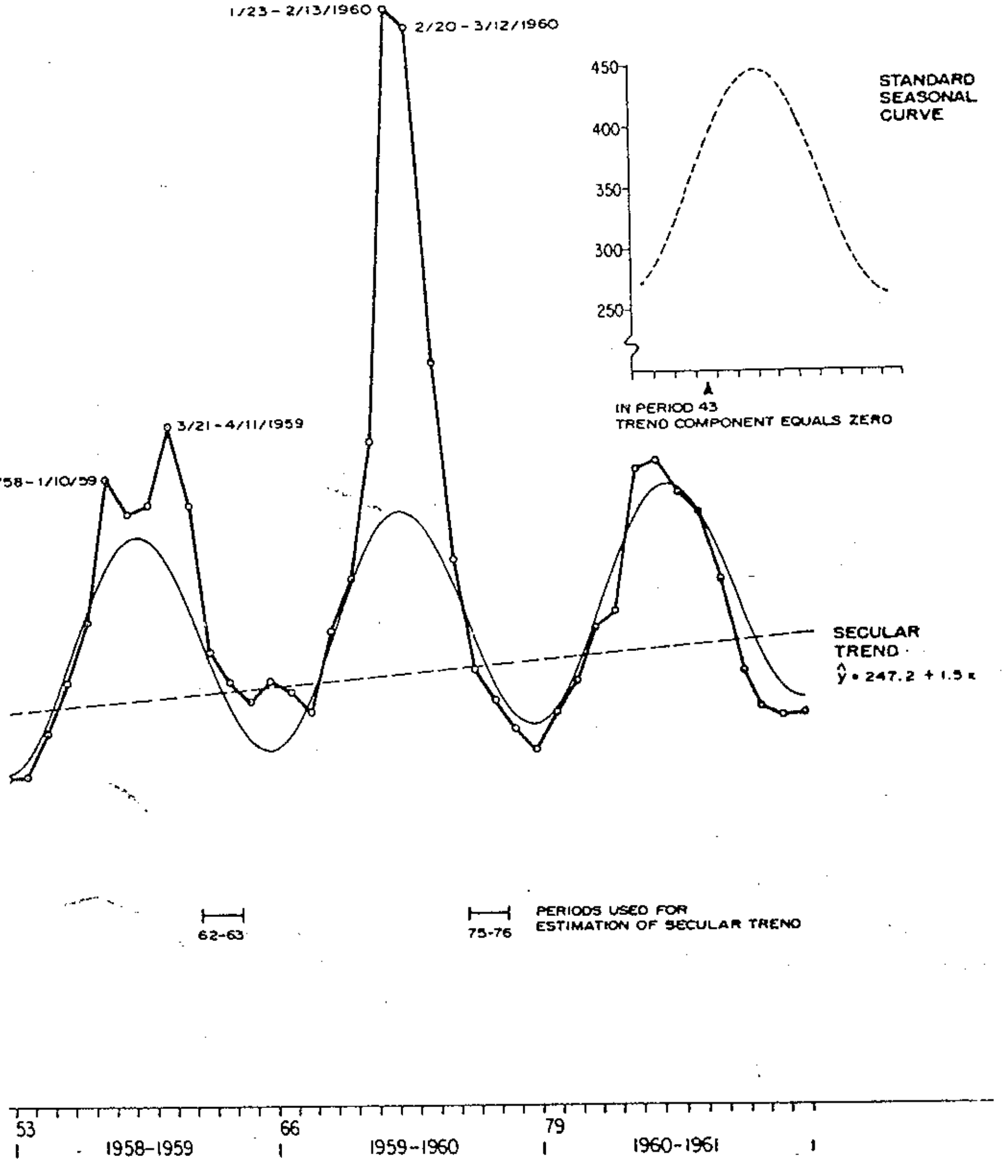


NOTE: Years by 4-week periods; first period of each year includes weeks 35–38, approximately weeks ending September 1 through September 22.

account for estimation of the three parameters of the sine function and the slope of the secular trend. A preliminary screening is advisable in order to remove the epidemic weeks and occasional very high or low values for single weeks which, as mentioned above, are commonly due to factors other than endemic variability. Taking 1958–59 and 1959–60 as base years, the average value of the standard deviation of deaths in the 108 cities for a single week was 27.9 deaths.

Discussion

The value of Selwyn Collins' methods for numerical description of influenza epidemics in the United States since the late 19th century is well known to epidemiologists and need not be reemphasized. The additional contribution of the present procedures is in the use of excess pneumonia-influenza mortality to provide current information. The comparison of this aspect of the present method with Collins' work is analogous to a comparison of the use of of pneumonia-influenza deaths in 108 U.S. cities



weekly reports of notifiable diseases for detection of communicable disease outbreaks with the analysis of annual reports that have been corrected for many defects in the hastily screened weekly data.

Current evaluation of weekly morbidity reports by the communicable disease control epidemiologist is based on experience with the characteristics of the reporting system for each disease integrated with medical knowledge, interpretation of laboratory data, and experience with past epidemics. In the same manner, weekly data on excess pneumonia-influenza mortality is one of many kinds of information which must be used in the early recognition of influenza epidemics and for week-by-week appraisal of geographic localization and relative magnitude. For this purpose the usefulness of the present methods has been indicated in various publications (*5–9*).

Distinct from practical evaluation of the procedure through use in surveillance of influenza is the statistical definition of a criterion for early recognition of an influenza epidemic and description of its characteristics.

Criterion of epidemic increase. The expected number of deaths may be regarded as an average value about which random variation will occur from week to week during interepidemic periods. The standard deviation, a measure of the random variation, is calculated from the weekly differences between recorded and expected deaths during interepidemic periods. In order to use the standard deviation to define an epidemic threshold for distinguishing random variation from epidemic increase, it must be assumed that the distribution of deviations can be described by some probability function. In the present method the bell-shaped "normal" probability function was used.

Using this function, with mean equal to the expected number of deaths in a given week and the epidemic threshold as described above, approximately 5 percent of the weekly deviations from the expected number of deaths will exceed the epidemic threshold.

Thus, during the course of a year in which there were no influenza epidemics, there would be, on the average, 2 or 3 weeks in each year in which the epidemic threshold was exceeded merely through chance fluctuation. Similarly, using the shorter period of 26 weeks to represent the winter and spring seasons during which influenza epidemics usually occur, there would be, on the average, 1 or 2 weeks in which the epidemic threshold was exceeded. Furthermore, the chance of one or more instances in which the epidemic threshold would be exceeded equals 0.75. Thus, in a nonepidemic year the odds are 3 to 1 that the epidemic threshold will be exceeded at least once. (Using the simple model of independent repeated trials with probability 0.05 of success on each trial, the chance of one or more successes in 26 trials equals $1-(0.95)^{26}=0.75$.)

Hence the use of a single week in which recorded deaths exceed the epidemic threshold as an indication of possible epidemic increase would not be satisfactory since this criterion would be satisfied rather often as a result of chance variation. In order to reduce the likelihood of being thus misled, the criterion has been set at 2 successive weeks during which the epidemic threshold is exceeded.

Using the model of repeated independent trials, but with somewhat more elaborate calculations (*10*), it can be shown that in 26 successive independent trials with probability of 0.05 for a success on each trial, there is only a 1-in-10 chance that one or more runs of two successes will occur. With this criterion the likelihood of being misled by chance fluctuations is thus reduced substantially. The probabilities given are approximate since there is a small correlation between successive deviations, and normality assumptions are not satisfied exactly; also change in projected secular trend is not taken into account.

Risk of failure to recognize an epidemic increase. Characteristically in influenza epidemics there is an initial small rise in excess mortality which is not distinguishable from random fluctuations in preceding weeks. Only rarely does this first rise exceed the epidemic threshold. In the second and third weeks, excess mortality usually increases rapidly, the amount varying with the magnitude of the epidemic.

In order to estimate the risk of failure to recognize an epidemic in its early stages by use of the criterion given, it is necessary to construct a standardized epidemic curve. For this purpose data for the 108 U.S. cities during the large influenza A2 epidemic of 1959–60 and the moderate-sized influenza B epidemic of 1961–62 were selected. Data for each epidemic were plotted, and a smoothed epidemic curve line was drawn. Readings from the smoothed lines

Table 6. Excess mortality during hypothetical influenza epidemics, based on experience in 108 U.S. cities

Early stages of epidemic	Week above expected number	Excess mortality in standard deviation units	
		Large epidemics	Moderate epidemics
Initial rise (usually below epidemic threshold).	First	1.0	0.5
Rapid increase (usually exceeds epidemic threshold).	Second	4.5	3.0
	Third	8.0	4.5

were then converted to standard deviations and rounded to give the values presented in table 6.

The difference between each value in table 6 and the epidemic threshold (1.64) was then calculated and regarded as a standardized normal deviation for which the probability that recorded deaths would exceed the epidemic threshold is obtained from standard tables. The complement of the products of these probabilities for 2 successive weeks gives the chance of failing to classify an increase in excess mortality as possibly epidemic. These probabilities are:

	Size of epidemic	
Period of epidemic	*Moderate*	*Large*
End of 2d week	0.9	0.7
End of 3d week	.1	<.01

For a typical influenza epidemic in which the initial rise in excess mortality does not exceed the epidemic threshold, there is a large probability of failure to classify the increase in mortality as epidemic at the end of the second week but a relatively small probability of failing to do so by the end of the third week of increased mortality.

Summary

Excess pneumonia-influenza mortality is used by the Communicable Disease Center in influenza surveillance, and various statistical procedures are employed for early recognition of excess mortality. Criteria for epidemic increase in mortality have been defined. The statistical methods all depend on short-term extrapolation of secular trend and a standard curve of seasonal change. Separate estimation of secular trend and seasonal change is recommended. A simple numerical procedure is used for calculation of the standard seasonal curve.

TECHNICAL NOTE

Derivation of computing formulas of table 4

The estimation procedure consists of carrying out the same operations on the observed values $Y_h-\bar{Y}$ as were described in the text for the sine function $A\sin\left(\frac{2\pi(h-1)}{N}+\Phi\right)$

Proceeding as before but replacing the operation of integration by summation gives

$$\sum_1^h (Y_h-\bar{Y})=\frac{1}{N}\left[N\sum_1^h Y_h-h\sum_1^N Y_h\right] \quad \text{[A1]}$$

which has a mean value over the range $h=1$ through $h=N$ of

$$\frac{1}{N}\left[\sum_1^N\sum_1^h Y_h-\left(\frac{N+1}{2}\right)\sum_1^N Y_h\right] \quad \text{[A2]}$$

Subtracting the mean value for each term of equation A1 and again cumulating through h gives

$$\frac{1}{N}\left[N\sum_1^h\sum_1^h Y_h-h\sum_1^N\sum_1^h Y_h+\frac{h(N-h)}{2}\sum_1^N Y_h\right] \quad \text{[A3]}$$

which has a mean value of

$$\frac{1}{N}\left[\sum_1^N\sum_1^h\sum_1^h Y_h-\left(\frac{N+1}{2}\right)\sum_1^N\sum_1^h Y_h+\left(\frac{N^2-1}{12}\right)\sum_1^N Y_h\right] \quad \text{[A4]}$$

Substituting

$$\sum_1^N Y_h=T_1,\ \sum_1^N\sum_1^h Y_h=T_2,\ \sum_1^N\sum_1^h\sum_1^h Y_h=T_3,\ \sum_1^h\sum_1^h Y_h=H \quad \text{[A5]}$$

and setting $[A3]=S_h/2N$, and $[A4]=-C/2N$, gives, using [A5]

$$\frac{S_h}{2N}=\frac{1}{2N}\left[2NH-h(2T_2)+h(N-h)T_1\right] \quad \text{[A6]}$$

$$\frac{C}{2N}=-\frac{1}{2N}\left[2T_3-(N+1)T_2+\left(\frac{N^2-1}{6}\right)T_1\right] \quad \text{[A7]}$$

Equating the sum of the above equations to $-\frac{AN^2}{4\pi^2}\sin\left(\frac{2\pi h}{N}+\Phi\right)$ gives

$$\frac{S_h}{2N}+\frac{C}{2N}=-\frac{AN^2}{4\pi^2}\sin\left(\frac{2\pi h}{N}+\Phi\right) \quad \text{[A8]}$$

multiplying each side of [A8] by $-\frac{4\pi^2}{AN^2}$ and adding $\bar{Y}$ to each side gives the estimating equation

$$\hat{Y}_{h+1}=\bar{Y}+A\sin\left(\frac{2\pi h}{N}+\Phi\right)=\bar{Y}-\frac{2\pi^2}{N^3}(S_h+C) \quad \text{[A9]}$$

REFERENCES

(1) Pennell, M. Y., Woolsey, T. H., Trantham, K. S., and Lehmann, J. L.: Selwyn D. Collins' contributions to health statistics: a guide to his works. PHS Publication No. 503 (Public Health Monograph No. 62). U.S. Government Printing Office, Washington, D.C., 1960.

(2) Bliss, C. I.: Periodic regression in biology and climatology. Bulletin 615. The Connecticut Agricultural Experiment Station, New Haven, June 1958.

(3) Searcy, J. K., Hardison, C. H., and Langbein, W. B.: Double mass curves. Geological Survey Water Supply Paper 1541–B. U.S. Geological Survey, Washington, D.C., 1960.

(4) Powell, R. W.: Successive integration. Annals of Mathematical Statistics 1: 123–136, May 1930.

(5) Trotter, Y., Jr., et al.: Asian influenza in the United States, 1957–58. Am. J. Hyg. 70: 34–50, July 1959.

(6) Langmuir, A. D., Pizzi, M., Trotter, W. Y., and Dunn, F. L.: Asian influenza surveillance. Pub. Health Rep. 73: 114–120, February 1958.

(7) Eickhoff, T. C., and Robinson, R. Q.: Influenza surveillance in the United States, 1960. Pub. Health Rep. 76: 1099–1106, December 1961.

(8) Eickhoff, T. C., Sherman, I. L., and Serfling, R. E.: Observations on excess mortality associated with epidemic influenza. J.A.M.A. 176: 776–782, June 3, 1961.

(9) Dauer, C. C., and Serfling, R. E.: Mortality from influenza, 1957–58 and 1959–60. Am. Rev. Resp. Dis. 83: 15–28, February 1961, pt. 2.

(10) Feller, W.: An introduction to probability theory and its application. Ed. 2. John Wiley & Sons, New York, 1957, ch. 13, sec. 7.

Perrott New Chairman of PHR Board of Editors

George St.J. Perrott has been appointed chairman of the Board of Editors of *Public Health Reports* for a 4-year term beginning June 1963. He succeeds Dr. Ernest L. Stebbins, dean of the School of Hygiene and Public Health at Johns Hopkins University.

Mr. Perrott retired from the Public Health Service in 1958, where he had been chief of the Division of Public Health Methods for nearly 20 years. During most of that period, he was managing director of *Public Health Reports*. He served as chairman of the Committee on Publications which conceived the present form of *Public Health Reports*, assumed in 1952.

Since his retirement Mr. Perrott has been engaged in projects dealing with health studies of human populations and use of medical services.

One of his early activities with the Public Health Service was to direct the first National Health Survey in 1935–36. He also directed development of the present National Health Survey, established in 1956 and now a major component of the National Center for Health Statistics.

Working closely with the War Manpower Commission during World War II, the Division of Public Health Methods developed the criteria and formulas used in allocating medical personnel among military and civilian services. During this period, Mr. Perrott was chairman of the Public Health Service's Committee on Postwar Planning. After the war, the division and its chief expanded studies of medical manpower and medical services, including studies of the economics of medical, dental, nursing, and public health schools; characteristics and distribution of medical personnel; group practice; and chronic illness and health of older people.

Mr. Perrott received his M.A. degree from the University of North Dakota in 1915. After graduate work in physical chemistry at Princeton University from 1915 to 1917 and service in World War I, he was employed by the U.S. Bureau of Mines at its Pittsburgh Station, where he worked on health and safety in mining and metallurgy. He was director of the station from 1927 to 1931.