unimpaired. With these findings all the parts of the puzzle, L and S of vaccinia, seemed to fall in place, and now more than ever it seems reasonable to conclude that they are nothing more than different components of a single protein molecule.

Little or nothing is known about the antigen that gives rise to antibodies that neutralize the virus of vaccinia. Indeed, most workers have been unable to remove these antibodies from immune serum by means of adsorption with purified elementary bodies. On the other hand, Salaman believes that there is a union between elementary bodies of vaccinia and neutralizing antibodies and that if sufficient amounts of elementary bodies are used the neutralizing substances can be adsorbed from immune serum. An assessment of information regarding the antigen that incites the production of neutralizing antibodies and the manner in which such antibodies act reveals that much remains to be learned concerning this the most important of all subjects connected with immunity to vaccinia.

From my remarks regarding viruses in general and vaccine virus in particular, it should be evident that there is nothing peculiar about immunity in virus diseases. Principles that hold in other fields operate also in the virus domain. Furthermore, it should be obvious that generalizations about immunity in virus maladies can be made with no more assurance than about resistance to other types of infection. Immunological and serological phenomena in each virus disease present special problems that have to be met not through generalizations but by specific experiments.

NORMS OF GROWTH

By Professor EDWIN B. WILSON

DEPARTMENT OF VITAL STATISTICS, HARVARD SCHOOL OF PUBLIC HEALTH

IN the *Proceedings* of the National Academy of Sciences (Vol. 21, pp. 633–4, 1935) I gave average heights and weights of 275 school girls for consecutive ages 7 to 16 years inclusive with the correlations of the heights and of the weights in the different years based on the measurements obtained in Dr. W. F. Dearborn's growth study. As many persons knew that the study involved many more than 275 girls, some have wondered why I took only the 275 for whom records were available for each and every one of the ten years from 7 to 16.

The answer, unless I am mistaken, is to be had by considering the aims of a growth study. If we desire to establish stable norms of height and weight or of other measurements at different ages we should take, of course, large samples because the standard error σ of an average is that of the distribution divided by the square root of the number n in the sample. Thus at 13 the average height was 153.40 cm with a standard deviation of 7.42, which means 153.40 ± .45 for the average if based on only 275 girls, whereas if we had measurements of four times as many the standard deviation of the mean would be only .22. However, for such norms one need not trouble with the continuity involved in growth studies; one could make a cross-sectional survey involving a large number of persons at each of the different ages.

Growth, however, is a continuous process and the amount of growth between two given ages is measured by a difference or increment in the measures. If we have l girls of one age and m different girls of another age as in cross-sectional studies with means Z and W, respectively, for some measurement and with standard deviations σ_z and σ_w, the sampling error of the difference $W-Z$ would be

$$\sigma^2_{W\text{-}Z} = \frac{\sigma_w^2}{m} + \frac{\sigma_z^2}{l};$$

but if we had n girls of both ages, as in growth studies, and the averages were X and Y, the sampling error of the difference could be obtained directly from the differences $y-x$ or indirectly from the correlation coefficient r between corresponding values of x and y as

$$\sigma^2_{Y\text{-}X} = \frac{\sigma_x^2 + \sigma_y^2 - 2r\sigma_x\sigma_y}{n}.$$

In cases in which r is high this value may be much smaller than the former when the number of persons involved is about the same, or, to put it differently, the second value may be statistically as good from a relatively small number of individuals, as the first is from a much larger number.

For example, the lowest correlation of the heights in successive years was found to be $r = .96$. If we assume σ_x and σ_y nearly enough equal so that they may be put equal, and equal to σ_z, σ_w for corresponding ages, without serious error, and if we take $l = m = n$ the first formula gives $2\sigma_x^2/n$ and the second gives $2\sigma_x^2(1-r)/n$ or only .04 as much; to put it inversely, we should have to have $l = m = 25n$ to obtain from the first formula a sampling error as small as that obtainable from the second, or we should need nearly 6,400 girls taken at each of the years to give as good an estimate of average growth as we got from 275 taken at both years—provided we trust our statistical formula.[1]

[1] This proviso may seem odd. We have, however, to remember that statistical formulas are mathematical theorems proved on certain assumptions which may not hold

This raises the further question of how to combine the data in case there are, as there needs must be, irregular omissions in the data. It is a general rule of statistics that if we have two independent and consistent[2] estimates Q_1 and Q_2 of a quantity with two standard deviations σ_1 and σ_2, the weighted mean $Q = pQ_1 + (1-p)Q_2$ will have the smallest value of σ_Q when $p = \sigma_2^2/(\sigma_1^2 + \sigma_2^2)$ and $\sigma_Q^{-2} = \sigma_1^{-2} + \sigma_2^{-2}$. Hence, applied to the estimates of amount of growth $Y-X$ and $W-Z$, the best estimate would have the sampling error

$$G = \frac{\sigma^2_{W-Z}(Y-X) + \sigma^2_{Y-X}(W-Z)}{\sigma^2_{W-Z} + \sigma^2_{Y-X}}.$$

$$\frac{1}{\sigma^2_G} = \frac{n}{\sigma^2_x + \sigma^2_y - 2r\sigma_x\sigma_y} + \frac{1m}{l\sigma^2_w + m\sigma^2_z}$$

If we use for illustration the assumption $\sigma_x = \sigma_y = \sigma_w = \sigma_z$, $l = m = tn$,

$$\sigma^2_G = \frac{2\sigma^2_x(1-r)}{n} \cdot \frac{1}{1 + (1-r)t},$$

and it is clear that if r is large so that $1-r$ is small, t must be considerable before an appreciable reduction is made in σ_G.

It is well known in statistics that the sampling error of a quantity involves the method of estimating the quantity. Thus if a universe is symmetrical, its center may be estimated from a sample drawn from the universe by the mean of the sample or by its median or by its mode or by the mean of the least and of the greatest element in the sample, but the standard deviations of the four estimates will be different. So in the above problem if we should undertake to form from X and Z at the earlier age and from Y and W at the later age the overall age means

$$\frac{nX + lZ}{n + l} \text{ and } \frac{nY + mW}{n + m}$$

and their difference

$$G = \frac{nY}{n+m} - \frac{nX}{n+l} + \frac{mW}{n+m} - \frac{lZ}{n+l}$$

we could get σ^2_G as

$$\sigma^2_G = \frac{n\sigma^2_y}{(n+m)^2} + \frac{n\sigma^2_x}{(n+l)^2} - \frac{2nr\sigma_x\sigma_y}{(n+m)(n+l)} + \frac{l\sigma^2_z}{(n+l)^2} + \frac{m\sigma^2_w}{(n+m)^2},$$

but this would be a bad way to estimate G if r were large and l and m were not large compared with n. Indeed, if we take the simple illustrative case as before, $\sigma_x = \sigma_y = \sigma_z = \sigma_w$, $l = m = tn$ we have

$$\sigma^2_G = \frac{2\sigma^2_x(1-r)}{n(1+t)^2}\left[1 + \frac{t}{1-r}\right]$$

This is greater than if we had omitted altogether the extra observations which were not common to both years unless $t \geqq (2r-1)/(1-r)$. If $r = .96$ we should have 23 times as many non-common as common observations before we should be as well off using general means to estimate growth.

This discussion will show, it is hoped, how important it is when establishing norms for increments of growth (*i.e.*, of *growth*) to maintain throughout the study a discipline on the part of the students and of the studied which will bring about the maximum continuity of the record.

for the observations to which the formulas are applied. Thus if the theoretical sampling error of some quantity Q for samples of n items be σ, and if we take a considerable number of samples of n items we may find that the standard deviation of the values of Q observed in the different samples is considerably more than the theoretical value σ. If we evaluate the amount of growth by subtracting averages taken for two groups at each of two ages and also evaluate it by averaging the amount of growth between those ages for a single group measured at both ages, doing this a considerable number of times for different single groups on the one hand and for different pairs of groups on the other, we may well find that the variations observed are not those given by theory and further that they are not in the same ratio as that given by theory. It often takes extended experience to correct for such differences between theory and observation, but in the absence of such experience we have to make our estimates according to the theory.

[2] The qualification that the estimates have to be consistent is usually omitted. There are cases to be found in the literature where inconsistent estimates have been combined by the rules which I believe to be appropriate only for consistent estimates. Thus W. S. Eichelberger and Arthur Newton, "The Orbit of Neptune's Satellite and the Pole of Neptune's Equator," Astronomical Papers of the American Ephemeris, Vol. 9, Pt. 3, 1926, pp. 275–337, discuss on p. 329 the value of the reciprocal of the mass of Neptune, finding from reduction of the visual observations $19176 + 25$ and from reduction of the photographic observations 19655 ± 36. The difference is 479, which is many times as much as would be consistent with the indicated errors, yet they obtain 19331 ± 21 by combining the observations as if they were consistent, even reducing the estimated error of the combination in accordance with the rule. With the high standard in the reduction of observations set for the American Ephemeris and Nautical Almanac by Simon Newcomb over many years, I have to be somewhat hesitant in suggesting the above criticism, yet I must say that I have never seen any theory of least squares which seems to me to validate the process by which the final result 19331 ± 21 is obtained from its immediate antecedents.

OBITUARY

WILLIAM REES BREBNER ROBERTSON

1881–1941

W. R. B. ROBERTSON was born on May 31, 1881, and spent his early life on a farm at Manchester, Kansas. He died in Iowa City on March 15, 1941. He was one of C. E. McClung's eager students of cytology in the University of Kansas (A.B., 1906; A.M., 1907). He also studied with E. L. Mark, 1909–1912, in Harvard (Ph.D., 1915). He then spent the rest of his