On Simultaneous Confidence Intervals for Multinomial Proportions

LEO A. GOODMAN*
University of Chicago

In this article we present a method for obtaining simultaneous confidence intervals for the parameters of a multinomial distribution, and we compare this method with the one suggested recently by Quesenberry and Hurst (1964). For the usual probability levels, we find, for example, that the confidence intervals introduced here have the desirable property that they are shorter than the corresponding intervals obtained by the Quesenberry-Hurst method. We also present methods for obtaining simultaneous confidence intervals for the differences among the parameters of the multinomial distribution, and we compare these methods with the one suggested earlier by Gold (1963) for studying linear functions of the multinomial parameters. For the usual probability levels, we find that the confidence intervals introduced in the present article have the desirable property that they are shorter than the corresponding intervals obtained by the Gold method applied to the differences among the multinomial parameters. In addition, we show how the methods presented here for studying the differences among the multinomial parameters can be modified in order to obtain simultaneous confidence intervals for the *relative* differences among the multinomial parameters.

1. INTRODUCTION AND SUMMARY

Let $n_1, n_2, \cdots, n_k$ denote the observed cell frequencies in a sample of size N from a multinomial distribution, and let $\pi_1, \pi_2, \cdots, \pi_k$ denote the corresponding parameters of the distribution. In other words, let n_i denote the number of observations falling in the ith cell $(i = 1, 2, \cdots, k)$ in a sample from a multinomial population, and let π_i denote the probability that an observation will fall in the ith cell $(i = 1, 2, \cdots, k)$. In a recent article, Quesenberry and Hurst (1964) proposed the following large-sample simultaneous confidence intervals for the k parameters $\pi_1, \pi_2, \cdots, \pi_k$:

$$\hat{\pi}_i^- \leq \pi_i \leq \hat{\pi}_i^+ \qquad (i = 1, 2, \cdots, k) \tag{1}$$

where

$$\hat{\pi}_i^- = \{A + 2n_i - \{A[A + 4n_i(N - n_i)/N]\}^{\frac{1}{2}}\}/[2(N + A)],$$
$$\hat{\pi}_i^+ = \{A + 2n_i + \{A[A + 4n_i(N - n_i)/N]\}^{\frac{1}{2}}\}/[2(N + A)],$$

Received June 1964; revised Dec. 1964.
* This research was sponsored by the Army Research Office, Office of Naval Research, and Air Force Office of Scientific Research by Contract No. Nonr-2121(23), NR 342-043 and was supported in part by Research Grant No. NSF-GP609 from the Division of Mathematical, Physical and Engineering Sciences of the National Science Foundation.

and where A is the upper $\alpha \times 100$-th percentage point of the chi-square distribution with $k - 1$ degrees of freedom. When $N \rightarrow \infty$ the probability will be at least $1 - \alpha$ that all of the k confidence intervals (1) will include the corresponding true values of the parameters. These confidence intervals were recommended by Quesenberry and Hurst for the situation where the probability should be $1 - \alpha$ (or more) that the confidence statement about all of the π_i $(i = 1, 2, \cdots, k)$ is correct; i.e.; where the probability should be α (or less) that this confidence statement is incorrect.

In the present article we shall present simultaneous confidence intervals for the π_i $(i = 1, 2, \cdots, k)$ which are modifications of the confidence intervals (1). The modified confidence intervals will continue to meet the requirement that the corresponding confidence statement about all of the π_i $(i = 1, 2, \cdots, k)$ will be correct with probability $1 - \alpha$ (or more), but the modified intervals will be shorter than those given by (1), for the usual probability levels (e.g., $\alpha = .01$, $.05$, or even $.10$). To illustrate the difference between the intervals introduced in the present article and those proposed by Quesenberry and Hurst, we shall reanalyze the data given in their article. We find, for example, that for data of this kind (where $k = 10$), at the probability level used by Quesenberry and Hurst (viz., $\alpha = .10$), each of their ten confidence intervals could be reduced in length by approximately one-third of its original length (when $N \rightarrow \infty$), and the confidence statement obtained with the reduced intervals would continue to be correct with probability .90 or more. Of course, the longer confidence intervals reduce the probability of an incorrect confidence statement, but in this situation where a 10 percent probability would be satisfactory (i.e., $\alpha = .10$), we find that the probability is actually less than two-tenths of one percent (i.e., $\alpha = .002$) for the longer confidence intervals (when $N \rightarrow \infty$), thus indicating the desirability of a reduction in length.

The methods described herein for obtaining simultaneous confidence intervals for the k parameters π_i $(i = 1, 2, \cdots, k)$ can be extended to provide simultaneous confidence intervals for the $k(k - 1)/2$ differences $\pi_i - \pi_j$ $(i > j)$ or for any other contrasts among the π_i. We shall discuss the estimation of these differences in the present article, and shall compare the simultaneous confidence intervals obtained here with those presented earlier by Gold (1963) for estimating linear functions of the π_i. (In this case the $\pi_i - \pi_j$ $(i > j)$ are the linear functions of interest.) In addition, we shall show how the methods presented here can be modified in order to provide simultaneous confidence intervals for the relative differences among the multinomial parameters.

2. Shorter Confidence Intervals for the π_i

Although Quesenberry and Hurst (1964) did not derive their confidence intervals by this method, it is possible to show that the confidence limits $\hat{\pi}_i^-$ and $\hat{\pi}_i^+$ in (1) can be obtained simply as the two solutions of a quadratic equation in π_i, viz.,

$$(p_i - \pi_i)^2 = A\pi_i(1 - \pi_i)/N, \qquad (i = 1, 2, \cdots, k) \tag{2}$$

where $p_i = n_i/N$. Quesenberry and Hurst set A equal to the upper $\alpha \times 100$-th percentile of the chi-square distribution with $k - 1$ degrees of freedom. For $k = 2$, the probability is $1 - \alpha$ that the confidence statement obtained from

(1) and (2) will be correct (when $N \to \infty$); but for $k > 2$ the probability will be greater than $1 - \alpha$. For $k = 2$, the confidence statement obtained from (1) and (2) corresponds to the usual large-sample confidence interval for the parameter of a binomial distribution (see, e.g., Mood and Graybill, 1963, p. 263). For $k > 2$, we find that this confidence statement can be improved by replacing A in (1) and (2) by B, where B is the upper $(\alpha/k) \times 100$-th percentile of the chi-square distribution with one degree of freedom. When A is replaced by B in (1) and (2), the probability is α/k that the confidence interval thus obtained will be incorrect for a single given parameter, say, π_i ; and therefore the probability is α (or less) that at least one of the k confidence intervals ($i = 1, 2, \cdots, k$) will be incorrect. (For a derivation of the inequality used in the above argument see, e.g., Wilks, 1962, p. 291.) For $k > 2$, we find that $B < A$ at the usual probability levels; e.g., $\alpha = .01$, .05, or .10. (For each value of k, the upper bound for the value of α for which $B < A$ can be determined using the tabulated chi-square distribution, but we shall not go into these details here.) Note that B can be determined from the tabulated chi-square distribution or from the tabulated normal distribution, since B is the square of the upper $(\alpha/2k) \times 100$-th percentile of the normal distribution with zero mean and unit variance.

In addition to providing shorter confidence intervals than those given by Quesenberry and Hurst, the preceding remarks also provide us with a method for obtaining a more accurate bound for the probability that their confidence statement will be correct. If we let α' denote the probability that a chi-square variate with one degree of freedom will exceed A, then the probability is α' that the confidence interval obtained from (1) and (2) will be incorrect for a single given parameter, say, π_i ; and therefore the probability is $k\alpha'$ (or less) that at least one of the k confidence intervals ($i = 1, 2, \cdots, k$) will be incorrect. For $k > 2$, we find that $k\alpha' < \alpha$ at the usual probability levels. This indicates that the upper bound of α, which is given by Quesenberry and Hurst for the probability that their confidence statement is incorrect, should be replaced when $k > 2$ by the more accurate bound $k\alpha'$.

Note that the value of A is a constant in each of the k confidence intervals (1). A more general set of simultaneous confidence intervals is obtained by replacing A by B_i in the confidence interval for π_i ($i = 1, 2, \cdots, k$), where B_i is the upper $\beta_i \times 100$-th percentile of the chi-square distribution with one degree of freedom, and $\sum_{i=1}^{k} \beta_i = \alpha$. If we consider the special case where $\beta_i = \alpha/k$ and $B_i = B$, we obtain the simultaneous confidence intervals introduced earlier herein. Both the special case and the more general method introduced above will be found useful in practice.

3. An Example

We shall now reanalyze the data given earlier by Quesenberry and Hurst. These data describe the frequency of ten different modes of failure as recorded in a study of 870 machines that failed.

For this example, Quesenberry and Hurst took $\alpha = .10$ and $A = 14.6837$ ($k = 10$). For the corresponding method introduced in the present article, we obtain $B = 6.6349$ when $\alpha = .10$ and $k = 10$. (We could replace the constant B by a set of ten constants B_i($i = 1, 2, \cdots, 10$) as described above, but we

TABLE 1

Frequency Distribution of Modes of Failure

Mode of Failure	1	2	3	4	5	6	7	8	9	10
Frequency	5	11	19	30	58	67	92	118	173	297

shall not go into these details here.) We now compare the Quesenberry-Hurst (Q-H) confidence intervals with the shorter intervals suggested here.

In view of the fact that $\alpha = .10$ is the upper bound for the probability that at least one of the ten shorter confidence intervals will be incorrect (when $N \to \infty$), the corresponding upper bound for the Q-H confidence intervals will be less than .10. Calculating the upper bound for the Q-H intervals by the method introduced in the preceding section, we find that this bound will actually be less than .002!

Comparing the shorter confidence intervals with the corresponding Q-H intervals with respect to their lengths, we find that the ratio of the two lengths vary, but we shall show in the following section that this ratio will converge in probability to $(B/A)^{\frac{1}{2}}$ when $N \to \infty$, which for the case under consideration here (where $k = 10$ and $\alpha = .10$) gives a value of approximately .67 for the ratio of the lengths.

4. Other Methods for Obtaining Confidence Intervals for the π_i

In the usual analysis of binomial proportions, the term $\pi_i(1 - \pi_i)$ in (2) would often be replaced by $p_i(1 - p_i)$ when N is large (see, e.g., Mood and Graybill, 1963, p. 263). If this replacement is now made in (2), this would then lead to the replacement of the simultaneous confidence intervals (1) by a simpler set of simultaneous confidence intervals; viz.,

$$\begin{aligned} \hat{\hat{\pi}}_i^- &= p_i - [Ap_i(1 - p_i)/N]^{\frac{1}{2}} \\ \hat{\hat{\pi}}_i^+ &= p_i + [Ap_i(1 - p_i)/N]^{\frac{1}{2}}, \end{aligned} \qquad (i = 1, 2, \cdots, k) \qquad (3)$$

TABLE 2

Confidence Intervals for the Mode-of-Failure Proportions

Mode of Failure	The Q-H Intervals $\hat{\pi}_i^-$	The Q-H Intervals $\hat{\pi}_i^+$	The Shorter Intervals $\tilde{\pi}_i^-$	The Shorter Intervals $\tilde{\pi}_i^+$
1	.001	.027	.002	.017
2	.004	.037	.006	.027
3	.009	.050	.012	.039
4	.017	.067	.022	.054
5	.041	.107	.048	.092
6	.049	.119	.057	.104
7	.072	.152	.082	.136
8	.097	.186	.108	.168
9	.152	.256	.166	.236
10	.283	.405	.301	.384

where A is as defined in Section 1 above. These approximate confidence intervals correspond to the simultaneous confidence intervals proposed earlier by Gold (1963) for linear functions of the π_i ($i = 1, 2, \cdots, k$), where in this case the π_i are the linear functions of interest. These confidence intervals are actually asymptotically equivalent to the corresponding Quesenberry-Hurst intervals (1). By an argument similar to that used in Section 2 above, we see that for the usual probability levels the confidence intervals (3) can be improved when $k > 2$ by replacing the A in (3) by B as defined in Section 2 above. Comparing the intervals obtained when B is used in (3) with the corresponding intervals obtained when A is used in (3), we find that the ratio of their lengths is simply $(B/A)^{\frac{1}{2}}$.

This result concerning the ratio of the lengths of the confidence intervals can also be applied to the Quesenberry-Hurst intervals (1) since they are asymptotically equivalent to the confidence intervals obtained when A is used in (3). Thus, comparing the intervals obtained when B is used in (1) with the corresponding intervals obtained when A is used in (1), we find that the ratio of their lengths will converge in probability to $(B/A)^{\frac{1}{2}}$ when $N \to \infty$.

Before closing this section, we take note of the fact that a simultaneous confidence statement about $k - 1$ of the parameters, say $\pi_1, \pi_2, \cdots, \pi_{k-1}$, will also provide us with a statement about the remaining parameter, say π_k, since $\pi_k = 1 - \sum_{i=1}^{k-1} \pi_i$. In particular, if the confidence intervals (3) are calculated for $i = 1, 2, \cdots, k - 1$, this simultaneous confidence statement implies that

$$1 - \sum_{i=1}^{k-1} \hat{\hat{\pi}}_i^+ \leq \pi_k \leq 1 - \sum_{i=1}^{k-1} \hat{\hat{\pi}}_i^- . \tag{4}$$

This then provides us with a confidence interval for π_k which is different from the corresponding interval calculated directly from (3); viz.,

$$\hat{\hat{\pi}}_k^- \leq \pi_k \leq \hat{\hat{\pi}}_k^+ . \tag{5}$$

However, it can be shown that

$$1 - \sum_{i=1}^{k-1} \hat{\hat{\pi}}_i^+ \leq \hat{\hat{\pi}}_k^-$$

and

$$1 - \sum_{i=1}^{k-1} \hat{\hat{\pi}}_i^- \geq \hat{\hat{\pi}}_k^+ , \tag{6}$$

and therefore the confidence interval (5) is shorter than the interval (4). Thus, the simultaneous confidence statement about the k parameters π_i ($i = 1, 2, \cdots, k$) that is obtained by calculating (3) for $i = 1, 2, \cdots, k$ is preferable to the statement obtained by calculating (3) for $i = 1, 2, \cdots, k - 1$ and (4). A similar result can be applied to the case where the A in (3) is replaced by B in all calculations.

5. Simultaneous Confidence Intervals for $\pi_i - \pi_j$

The methods proposed above can be extended to provide simultaneous confidence intervals for the $k(k - 1)/2$ differences $\Delta_{ij} = \pi_i - \pi_j$ ($i > j$). In this case the method proposed by Gold (1963) would yield the K intervals

$$\hat{\Delta}_{ij}^{-} = d_{ij} - [A(p_i + p_j - d_{ij}^2)/N]^{\frac{1}{2}} \qquad (i > j) \qquad (7)$$
$$\hat{\Delta}_{ij}^{+} = d_{ij} + [A(p_i + p_j - d_{ij}^2)/N]^{\frac{1}{2}},$$

where $d_{ij} = p_i - p_j$, $K = k(k - 1)/2$, and A is as defined in Section 1 above. By an argument similar to that used in Section 2 above, we find that for the usual probability levels these simultaneous confidence intervals can be improved when $k > 2$ by replacing the A in (7) by C, where C is the upper $(\alpha/K) \times 100$-th percentile of the chi-square distribution with one degree of freedom. (If the set of simultaneous confidence intervals includes the k intervals (3) for the π_i $(i = 1, 2, \cdots, k)$ and also the K intervals (7) for the Δ_{ij} $(i > j)$, then for the usual probability levels an improvement can be obtained in both (3) and (7) when $k > 4$ by replacing the A in (3) and (7) by D, where D is the upper $(\alpha/L) \times 100$-th percentile of the chi-square distribution with one degree of freedom, and $L = k + K = k(k + 1)/2$.)

The K differences Δ_{ij} $(i > j)$ are particular examples of linear functions of the π_i. In the situation where we may be interested in all possible linear functions of the π_i (i.e., functions of the form $\Delta(\underset{\sim}{a}) = \sum_{i=1}^{k} a_i \pi_i$), the method proposed by Gold would yield the following set of simultaneous confidence intervals for the $\Delta(\underset{\sim}{a})$:

$$\hat{\Delta}(\underset{\sim}{a}) = d(\underset{\sim}{a}) \pm [AS^2(\underset{\sim}{a})/N]^{\frac{1}{2}}, \qquad (8)$$

where $d(a) = \sum_{i=1}^{k} a_i p_i$ and

$$S^2(\underset{\sim}{a}) = \sum_{i=1} a_i^2 p_i - [d(\underset{\sim}{a})]^2. \qquad (9)$$

If a particular set of M linear functions are of interest, say $\Delta_1, \Delta_2, \cdots, \Delta_M$, the methods used in Section 2 above can be extended to show that simultaneous confidence intervals for these linear functions can be calculated from (8) applied to the M linear functions with A in (8) replaced by E, where E is the upper $(\alpha/M) \times 100$-th percentile of the chi-square distribution with one degree of freedom. We noted earlier in this article that if $M = k$ or $M = k(k - 1)/2$ then $E < A$ for the usual probability levels when $k > 2$. For any particular application in which the probability level and the value of M have been specified, calculation of both E and A would indicate which should be used in (8). To reduce the length of the confidence intervals, replace A by E whenever $E < A$.

6. Simultaneous Confidence Intervals for π_i/π_j

The methods presented above for studying $\pi_i - \pi_j$ $(i > j)$ can be extended to provide similar methods for studying π_i/π_j $(i > j)$ or $\log(\pi_i/\pi_j)$, the natural logarithm of π_i/π_j, which we denote by β_{ij}. The simultaneous confidence intervals for the $\pi_i - \pi_j$ were obtained by considering the distribution of the estimators $p_i - p_j$, and similarly simultaneous confidence intervals for the π_i/π_j or for the β_{ij} can be obtained by considering the distribution of the corresponding estimators p_i/p_j or $b_{ij} = \log(p_i/p_j)$. The b_{ij} are particular examples of contrasts of the logarithms (i.e., $b_{ij} = \log p_i - \log p_j$), and we can determine the distribution of the b_{ij} by regarding the $\log p_i$ and $\log p_j$ as asymptotically independent normal variates with variance estimated by n_i^{-1} and n_j^{-1}, respectively (see Plackett, 1962, and Goodman, 1963a). Thus, the variance of b_{ij} can be estimated by

$$S^2(b_{ij}) = n_i^{-1} + n_j^{-1} . \tag{10}$$

By the application of methods similar to those used earlier by the author (1964a, p. 93), we obtain the following simultaneous confidence intervals for the $k(k-1)/2$ values of β_{ij} $(i > j)$:

$$\begin{aligned} \hat{\beta}_{ij}^- &= b_{ij} - [AS^2(b_{ij})]^{\frac{1}{2}} \\ \hat{\beta}_{ij}^+ &= b_{ij} + [AS^2(b_{ij})]^{\frac{1}{2}}, \end{aligned} \tag{11}$$

where $S^2(b_{ij})$ is calculated by (10), and A is (as above) the upper $\alpha \times 100$-th percentile of the chi-square distribution with $k - 1$ degrees of freedom.

We noted above that the β_{ij} are particular examples of contrasts of the $\log \pi_i$. In the situation where we may be interested in all possible contrasts of the $\log \pi_i$ (i.e., in all functions of the form $\beta(\underset{\sim}{a}) = \sum_{i=1}^k a_i \log \pi_i$, where $\sum_{i=1}^k a_i = 0$), the method given above can be extended to provide the following set of simultaneous confidence intervals for the $\beta(\underset{\sim}{a})$:

$$\hat{\beta}(\underset{\sim}{a}) = b(\underset{\sim}{a}) \pm [AS^2(b(\underset{\sim}{a}))]^{\frac{1}{2}}, \tag{12}$$

where

$$b(\underset{\sim}{a}) = \sum_{i=1}^{k} a_i \log p_i \quad \text{and} \quad S^2(b(\underset{\sim}{a})) = \sum_{i=1}^{k} a_i^2 n_i^{-1} . \tag{13}$$

If a particular set of M contrasts are of interest, say $\beta_1, \beta_2, \cdots, \beta_M$, the methods used in Section 2 above can be applied to show that simultaneous confidence intervals for these contrasts can be calculated from (12) applied to the M linear functions with A in (12) replaced by E, where E is the upper $(\alpha/M) \times 100$-th percentile of the chi-square distribution with one degree of freedom. (See related comments in Section 5.) In particular, if the $k(k-1)/2$ values of β_{ij} are of interest, then $M = k(k-1)/2$ and $E < A$ for the usual probability levels when $k > 2$. Thus, in this case the confidence intervals (11) and (12) can be improved by replacing A in (11) and (12) by E.

The confidence intervals presented above for the β_{ij} and $\beta(\underset{\sim}{a})$ can also be applied to obtain confidence intervals for the ratios π_i/π_j, which we now denote by γ_{ij}. Note that $\gamma_{ij} = \exp[\beta_{ij}]$. From (11) we obtain the following simultaneous confidence intervals for the $k(k-1)/2$ values of γ_{ij} $(i > j)$:

$$\begin{aligned} \hat{\gamma}_{ij}^- &= \exp[\hat{\beta}_{ij}^-] = (p_i/p_j)/\exp\{[AS^2(b_{ij})]^{\frac{1}{2}}\} \\ \hat{\gamma}_{ij}^+ &= \exp[\hat{\beta}_{ij}^+] = (p_i/p_j)\exp\{[AS^2(b_{ij})]^{\frac{1}{2}}\}. \end{aligned} \tag{14}$$

A similar modification of (12) will provide simultaneous confidence intervals for all functions of the form

$$\gamma(\underset{\sim}{a}) = \prod_{i=1}^{k} \pi_i^{a_i} , \tag{15}$$

where $\sum_{i=1}^k a_i = 0$, and the remarks in the preceding paragraph concerning the replacement of A by E will apply here as well.

7. Further Remarks

In the present article we have been concerned with the estimation of the parameters of a single multinomial population. The ideas and methods developed here can be extended to situations where comparisons between two (or more) multinomial populations are of interest. For example, in situations where differences among multinomial parameters are of interest, the differences between the corresponding parameters of two (or more) multinomial populations can be analyzed by methods that are closely related to, but different in detail from, the methods presented in Section 5 above (see Goodman, 1964c). In situations where relative differences (or ratios) among multinomial parameters are of interest, the relative differences between the corresponding ratios in two (or more) multinomial populations can be analyzed by methods that are closely related to, but different in detail from, the methods presented in Section 6 above (see Goodman, 1964a). In addition, two (or more) sets of two binomial populations can be compared in situations where differences between the parameters are of interest using the generalization of the Stouffer method given by the author in (1963b), and two (or more) sets of J multinomial populations can be compared in situations where ratios of the parameters are of interest using the various methods that have been developed for the analysis of three-factor interaction in a three-way contingency table (see, e.g., Goodman, 1964b and the literature cited there).

Although the methods developed in the present article are different from the usual methods suggested for the analysis of multinomial data, they all have in common the feature that they are derived without recourse to a prior distribution of the parameters of the multinomial distribution. For a treatment of this subject when certain special prior distributions are used, we refer the reader to a recent paper by Lindley (1964).

References

Gold, R. Z. 1963. Tests auxiliary to χ^2 tests in a Markov chain. *Ann. Math. Statist.*, *34*, 56–74

Goodman, L. A. 1963a. On Plackett's test for contingency table interactions. *J. Roy. Statist. Soc. Ser. B*, *25*, 179–188.

Goodman, L. A. 1963b. On methods for comparing contingency tables. *J. Roy. Statist. Soc. Ser. A*, *126*, 94–108.

Goodman, L. A. 1964a. Simultaneous confidence limits for cross-product ratios in contingency tables. *J. Roy. Statist. Soc. Ser. B*, *26*, 86–102.

Goodman, L. A. 1964b. Simple methods for analyzing three-factor interaction in contingency tables. *J. Amer. Statist. Assoc.*, *59*, 319–352.

Goodman, L. A. 1964c. Simultaneous confidence intervals for contrasts among multinomial populations. *Ann. Math. Statist.*, *35*, 716–725.

Lindley, D. V. 1964. The Bayesian analysis of contingency tables. *Ann. Math. Statist.*, *35*, 1622–1643.

Mood, A. M. and Graybill, F. A. 1963. *Introduction to the Theory of Statistics*, Second Edition, McGraw-Hill, New York.

Plackett, R. L. 1962. A note on interactions in contingency tables. *J. Roy. Statist. Soc. Ser. B*, *24*, 162–166.

Quesenberry, C. P., and Hurst, D. C. 1964. Large sample simultaneous confidence intervals for multinomial proportions. *Technometrics*, *6*, 191–5.

Scheffé, H. 1959. *The Analysis of Variance*, Wiley, New York.

Tukey, J. W. 1953. The problem of multiple comparisons. Dittoed MS, Princeton Universitv.

Wilks, S. S. 1962. *Mathematical Statistics*, Wiley, New York.