United States
Environmental Protection
Agency

Health Effects Research
Laboratory
Research Triangle Park NC 27711

EPA-600/9-80-029
May 1980

Research and Development

EPA

Small Sample Properties of Various Tests on 2x2 Tables

RESEARCH REPORTING SERIES

Research reports of the Office of Research and Development, U.S. Environmental Protection Agency, have been grouped into nine series. These nine broad categories were established to facilitate further development and application of environmental technology. Elimination of traditional grouping was consciously planned to foster technology transfer and a maximum interface in related fields. The nine series are:

1. Environmental Health Effects Research
2. Environmental Protection Technology
3. Ecological Research
4. Environmental Monitoring
5. Socioeconomic Environmental Studies
6. Scientific and Technical Assessment Reports (STAR)
7. Interagency Energy-Environment Research and Development
8. "Special" Reports
9. Miscellaneous Reports

This document is available to the public through the National Technical Information Service, Springfield, Virginia 22161.

EPA-600/9-80-029
May 1980

SMALL SAMPLE PROPERTIES OF VARIOUS TESTS ON 2 x 2 TABLES

by

Victor Hasselblad
Andrew G. Stead
Statistics and Data Management Office
Health Effects Research Laboratory
Research Triangle Park, North Carolina 27711

Health Effects Research Laboratory
Office of Research and Development
U.S. Environmental Protection Agency
Research Triangle Park, North Carolina 27711

DISCLAIMER

This report has been reviewed by the Health Effects Research Laboratory, U.S. Environmental Protection Agency, and approved for publication. Mention of trade names or commercial products does not constitute endorsement or recommendation for use.

FOREWARD

The many benefits of our modern, developing, industrial society are accompanied by certain hazards. Careful assessment of the relative risk of existing and new man-made environmental hazards is necessary for the establishment of sound regulatory policy. These regulations serve to enhance the quality of our environment in order to promote the public health and welfare and the productive capacity of our nation's population.

The Health Effects Research Laboratory, Research Triangle Park, conducts a coordinated environmental health research program in toxicology, epidemiology, and clinical studies using human volunteer subjects. These studies address problems in air pollution, non-ionizing radiation, environmental carcinogenesis and the toxicology of pesticides as well as other chemical pollutants. The Laboratory participates in the development and revision of air quality criteria documents on pollutants for which national ambient air quality standards exist or are proposed, provides the data for registration of new pesticides or proposed suspension of those already in use, conducts research on hazardous and toxic materials, and is primarily responsible for providing the health basis for non-ionizing radiation standards. Direct support to the regulatory function of the Agency is provided in the form of expert testimony and preparation of affidavits as well as expert advice to the Administrator to assure the adequacy of health care and surveillance of persons having suffered imminent and substantial endangerment of their health.

In this paper, the small sample properties of the analysis of categorical data methods of Grizzle, Starner, and Koch are explored. Also, the small sample properties of maximum likelihood techniques are examined. Both analysis techniques have been used to analyze cross-sectional studies of respiratory behavior, and an understanding of the behavior of these tests when applied to small samples will enhance their applicability in future respiratory studies.

F. G. Hueter, Ph. D.
Director
Health Effects Reserach Laboratory

ABSTRACT

With particular interest in the likelihood ratio tests and tests generated by the categorical analysis methods of Grizzle, Starmer, and Koch, the small sample behavior of actual α levels is graphically illustrated for several asymptotically chi-square statistics when testing is performed on 2 x 2 contingency table data at a nominal level of .05. This behavior is examined for hypotheses testing both homogeneity and independence. In the homogeneity case, the effects on these levels are examined for equal and unequal sample sizes. In the independence case, the effects of different constraints concerning individual cell probabilities are demonstrated. Visual comparisons of the power of the likelihood ratio and the Grizzle, Starmer, Koch related statistics are also provided.

SECTION 1

INTRODUCTION

In several cross-sectional studies of respiratory conditions conducted by the U. S. Environmental Protection Agency [1-3], the resulting multiway contingency tables were analyzed using the analysis of categorical data methods of Grizzle, Starmer, and Koch (GSK). As these authors indicate: "the behavior of the tests in small samples is unknown" [4].

Additionally, since it is often possible to analyze the same date using maximum likelihood techniques, we were quite interested in examining the small sample properties of both procedures from the hypothesis testing standpoint.

Although simulation could be used to study the properties of these tests in larger tables, we restricted our investigation to the 2 x 2 table. In this way, the exact probability levels of the tests could be calculated.

Literature discussing various aspects of 2 x 2 tables abounds. Several papers, however, are particularly pertinent to our study. In 1967, Grizzle [5] investigated the effect of the continuity correction when testing for either homogeneity or independence. Garside and Mack [6], in 1976, discussed actual Type I error probabilities for various tests of the homogeneity case. In 1977, Eberhardt and Fligner [7] compared actual α levels, as well as the power, of the usual χ^2 test for homogeneity with a test equivalent to Goodman's γ^2 in small to moderate samples. More recently (1978), Larntz [8] compared exact levels of the likelihood ratio, Freeman-Tukey, and Pearson chi-squared goodness-of-fit statistics in small samples.

Our paper will compare the following tests, when appropriate, under both hypotheses described by Grizzle [5]: 1) the GSK, assuming either a linear or log-linear model, 2) the likelihood ratio test, 3) the usual χ^2 test, 4) the χ^2 test with Yates continuity correction, and 5) Fisher's "exact" test. The GSK test in the homogeneity case is the same as the γ^2 statistic proposed by Goodman [9].

SECTION 2

THE ASSUMPTIONS AND TESTS

A sample of n observations has been classified according to two dichotomous characteristics, A and B. Using the notation of Grizzle [5], we have the following table:

	A_1	A_2	
B_1	X_{11}	X_{12}	$X_{1.}$
B_2	X_{21}	X_{22}	$X_{2.}$
	$X_{.1}$	$X_{.2}$	n

The two sampling methods of interest are: 1) $X_{.1}$ and $X_{.2}$ fixed, and the hypotheses compare the difference in proportions of two binomial polulations, i.e. H_0: $p_1 = p_2$; and 2) only n is fixed, and the sample follows a multinomial distribution, and H_0: $P_{ij} = p_{i.}p_{.j}$ for i,j = 1,2.

In both cases it is possible to generate all possible samples and their associated probabilities for 1) fixed $p_{i.}$'s or 2) fixed p_{ij}'s. In this way actual α levels can be computed for small samples. Ahtough theoretically possible, calculating such levels for larger tables would require prohibitive amounts of computer time.

<u>Case I</u> - H_0: $p_1 = p_2$

The following test statistics were considered:

1) The uncorrected chi-square

$$X_u^2 = n(X_{11}X_{22}-X_{21}X_{12})^2/(X_{1.}X_{2.}X_{.1}X_{.2}), \qquad (2.1)$$

2) The corrected chi-square suggested by Yates

$$X_c^2 = n(|X_{11}X_{22}-X_{21}X_{12}|-n/2)^2/(X_{1.}X_{1.}X_{.1}X_{.2}), \qquad (2.2)$$

3) The GSK method[1] for testing the linear difference in two proportions,

$$X_{GSK}^2 = (X_{.2}X_{11}-X_{.1}X_{12})^2X_{.1}X_{.2}/(X_{11}^3X_{12}X_{22}+X_{12}^3X_{11}X_{21}) \qquad (2.3)$$

and,

4) The likelihood ratio test,

$$X^2_{LR} = -2(X_{1\cdot}\ln(X_{2\cdot}/n)+X_{2\cdot}\ln(X_{2\cdot}/n)-X_{11}\ln(X_{11}/X_{\cdot 1})$$
$$-X_{21}\ln(X_{21}/X_{\cdot 1})-X_{12}\ln(X_{12}/X_{\cdot 2})-X_{22}\ln(X_{22}/X_{\cdot 2})). \qquad (2.4)$$

Since these statistics are asymptotically chi-square, they were tested against the appropriate χ^2 value with one degree of freedom with $\alpha = .05$.

<u>Case II</u> - H_0: $p_{ij} = p_{i\cdot}p_{\cdot j}$, for $i,j = 1,2$

For the multinomial or independence case, five test statistics were examined:

1) The uncorrected X^2 as in Case I,

2) The corrected X^2 as in Case I,

3) The GSK method[1] for testing independence using a log-linear model

$$X^2_{GSK} = [\ln(X_{11}/n)-\ln(X_{12}/n)-\ln(X_{21}/n)+\ln(X_{22}/n)]^2(\sum_{ij}\sum\frac{1}{X_{ij}}), \qquad (2.5)$$

4) The likelihood ratio test

$$X^2_{LR} = 2\ [\sum_{ij}\sum X_{ij}\ln(X_{ij})-n_{1\cdot}\ln(n_{1\cdot})-n_{2\cdot}\ln(n_{2\cdot})$$
$$-n_{1}\ln(n_{\cdot 1})-n_{\cdot 2}\ln(n_{\cdot 2})] \qquad (2.6)$$

5) Fisher's "exact" test as defined in Kendall and Stuart [10],

$$p = \sum_{X=0}^{X_{11}} \binom{X_{1\cdot}}{X}\binom{X_{2\cdot}}{X_{21}}/\binom{n}{X_{\cdot 1}} \qquad (2.7)$$

As in Case I, all tests, except Fisher's "exact," are asymptotically chi-square with one degree of freedom. As randomization was not used, <u>Fisher's "exact" test will</u> necessarily be quite conservative.

[1]If any of the X_{ij}'s were 0, they were replaced by one over the number of response categories as suggested by the authors.

SECTION 3

RESULTS

Figures 1 and 2 depict the actual α levels obtained for the four tests considered in Case I. Behavior of the actual α levels for four of the five tests of Case II is shown in Figures 3 and 4. All the tests are for a nominal α level of .05. Figure 5 compares the α level performance of the corrected chi-square with Fisher's "exact" test. Additionally, some power curve comparisons between the GSK and likelihood ratio tests appear in Figures 6A-6D.

Figures 1A-1D show how the actual α levels of the four tests change for total sample sizes of 20, 40, 60 and 80 when testing the difference in two proportions by sampling in equal numbers from two binomial populations. Corresponding to Figures 1A-1D, Table 1 gives the maximum α level, over the entire range of p, for each test statistic and sample size.

The uncorrected chi-square statistic (Fig. 1A) had α levels very close to .05, and although it behaved better for smaller samples, actual alpha levels never exceeded .0590. The corrected chi-square statistic (Fig. 1B) produced very conservative α levels. Even with a total sample size of 80, its maximum alpha level was only .0330.

For total samples of 20 and 40, the GSK statistic (Fig. 1C) had very high levels. For a total sample of size 60, however, its maximum α level dropped to .0602. At 80, it appeared to be as good as the uncorrected chi-square statistic. The likelihood ratio test statistic (Fig. 1D) produced high maximum α levels for all sample sizes. As the size of sample increased, this problem occurred mainly for values of p_1 ($=p_2$) near 0 or 1.

In Figures 2A-2D, we examined the same test statistics and total sample sizes as in Figures 1A-1D, but sampled in a 3 to 1 ratio from the two populations. This unequal sampling had the following effects on the four test statistics: 1) maximum α levels (Table 2) for the uncorrect chi-square were generally closer to the nominal level, 2) the corrected chi-square was more conservative, 3) the GSK and likelihood ratio generally produced higher α levels, with the GSK uniformly high for $.2 \leq p_1 \leq .8$.

For Case II (multinomial sampling), one can visually examine the behavior of the actual α levels of four of the five test statistics in Figures 3 and 4 for the same total sample sized used in Case I. Probability levels for Fisher's "exact" test will be discussed later.

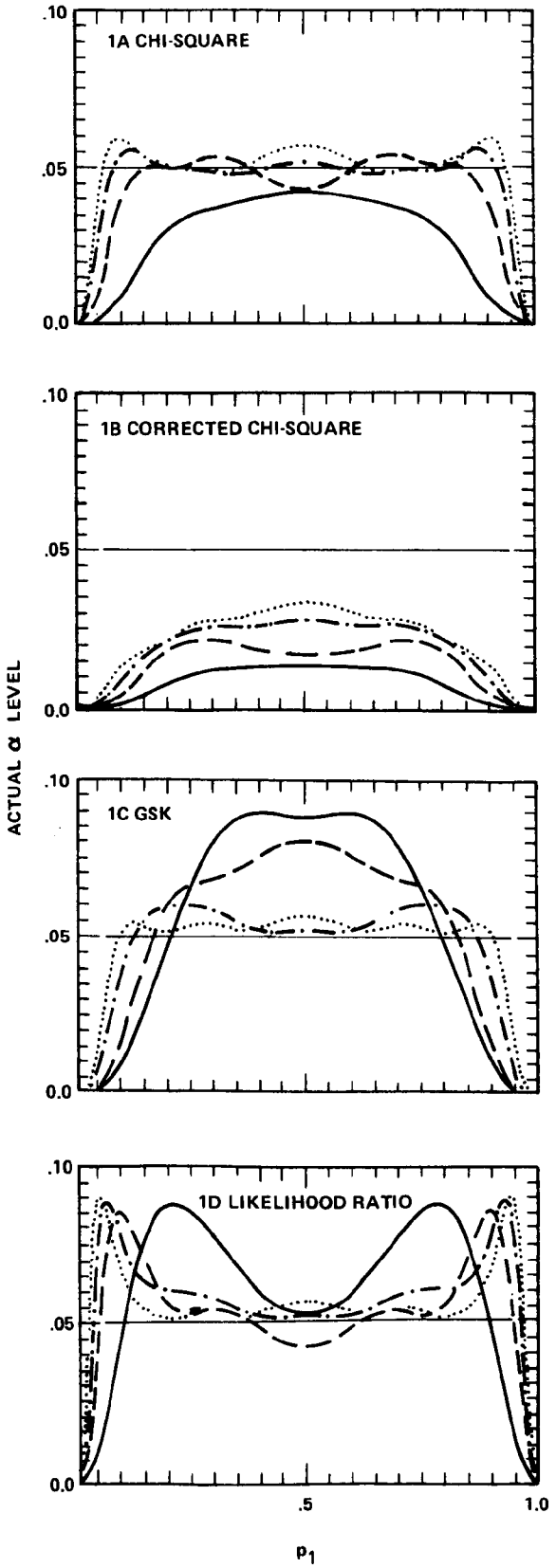

Figure 1. Actual α levels of four tests as a function of p_1 (=p_2), where $N_1 = N_2$.

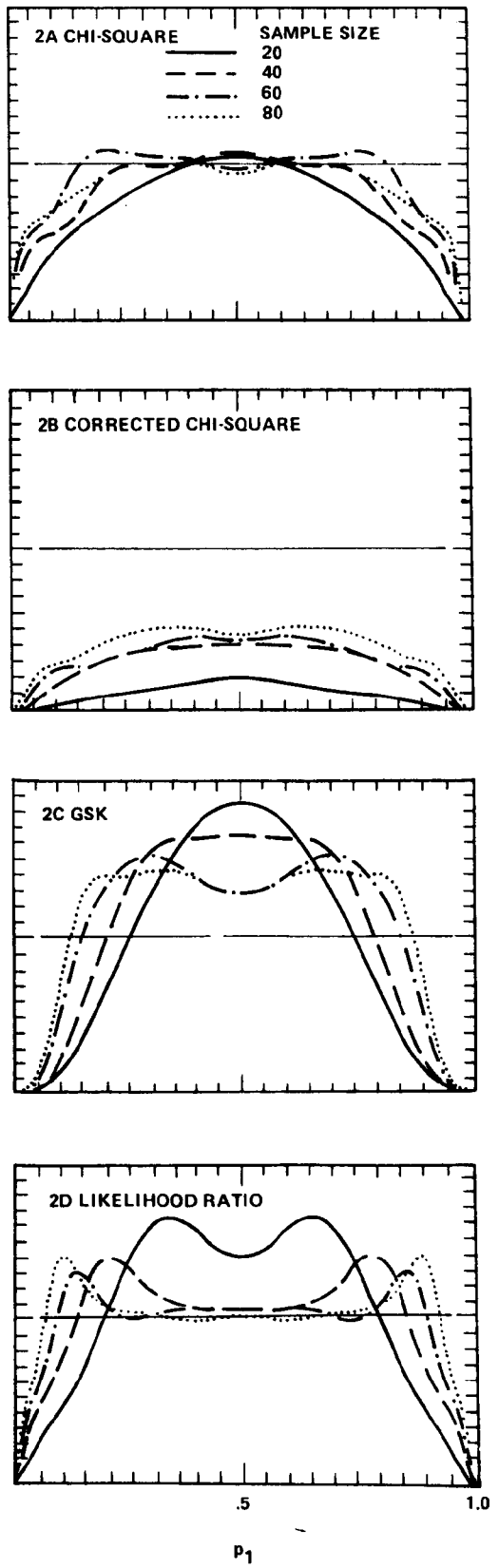

Figure 2. Actual α levels of four tests as a function of p_1 (=p_2), where $N_2 = 3N_1$.

TABLE 1.

MAXIMUM α LEVELS FOR TESTS OF DIFFERENCES IN PROPORTIONS OF TWO BINOMIALS OF EQUAL SIZE

Sample Size		Chi-square Uncorrected	Chi-square Corrected	GSK	Likelihood Ratio
$X_{.1}$	$X_{.2}$				
10	10	.0422	.0128	.0894	.0876
20	20	.0534	.0215	.0807	.0853
30	30	.0554	.0274	.0602	.0891
40	40	.0590	.0330	.0567	.0903

TABLE 2.

MAXIMUM α LEVELS FOR TESTS OF DIFFERENCES IN PROPORTIONS OF TWO BINOMIALS OF UNEQUAL SIZE

Sample Size		Chi-square Uncorrected	Chi-square Corrected	GSK	Likelihood Ratio
$X_{.1}$	$X_{.2}$				
5	15	.0521	.0106	.0930	.0856
10	30	.0542	.0211	.0827	.0733
15	45	.0546	.0235	.0768	.0687
20	60	.0509	.0269	.0715	.0732

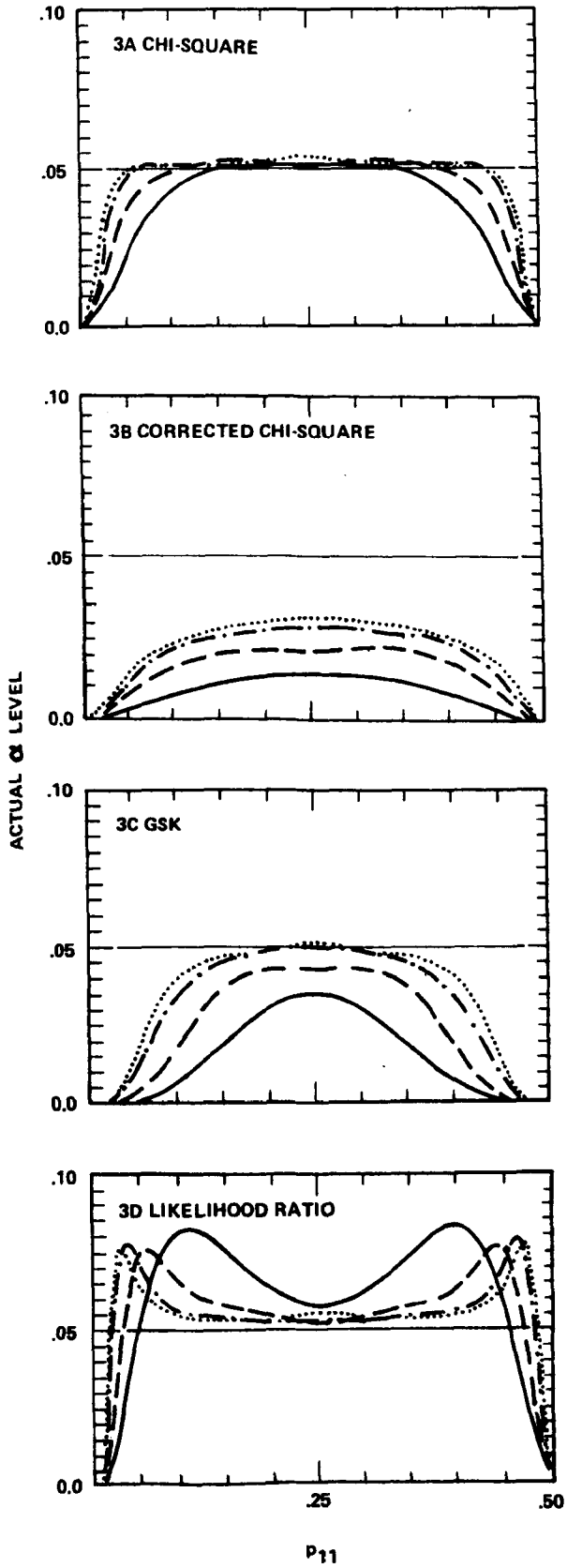

Figure 3. Actual α levels of four tests as a function of p_{11}, where $p_{12} = .5 - p_{11}$ and $p_{21} = p_{11}$.

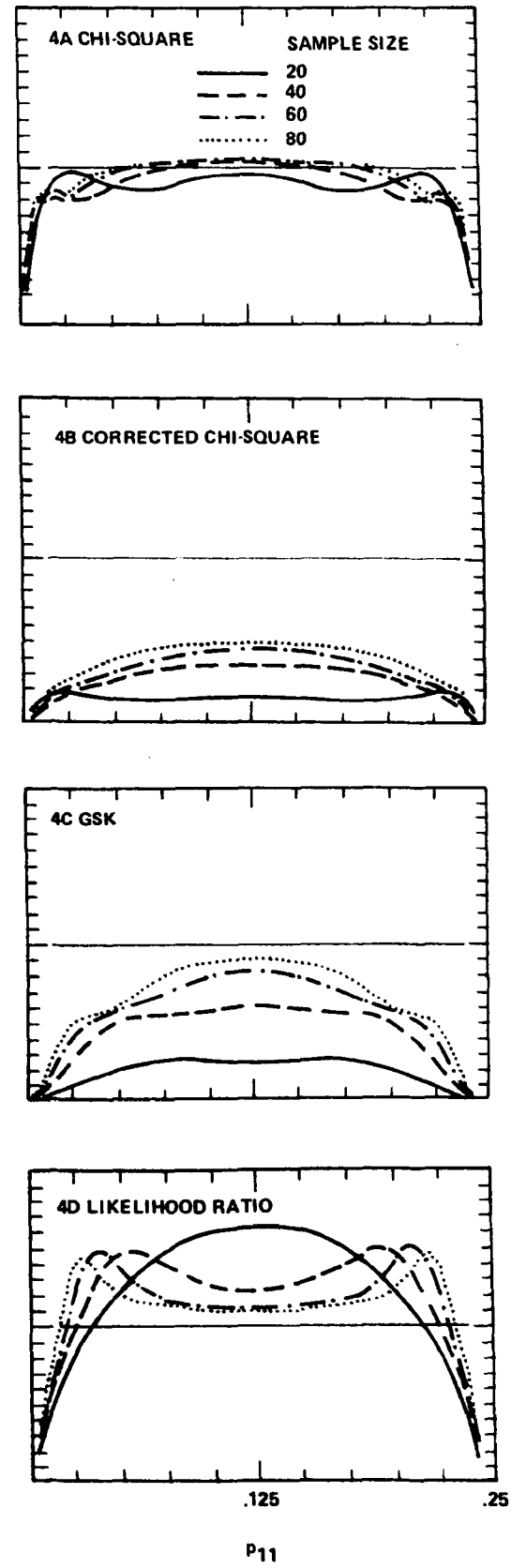

Figure 4. Actual α levels of four tests as a function of p_{11}, where $p_{12} = .25 - p_{11}$, $p_{21} = 3p_{11}$.

The plots in Figure 3 were constructed under the constraints that $p_{11} = p_{12}$, $p_{21} = p_{22}$, and $p_{11} + p_{21} = .5$. *Without loss of generality*, we then calculated actual levels for $0 < p_{11} < .5$. In Figure 4, a second set of constraints was imposed, i.e. $3\,p_{11} = p_{12}$, $3\,p_{21} = p_{22}$, and $p_{11} + p_{21} = .25$. In this instance, we computed actual α levels for $0 < p_{11} < .25$.

As in Case I, alpha levels of the uncorrected chi-square statistic adhere very closely to the nominal .05 level of the test for all sample sizes. By increasing the size of the sample, we observe (Figure 3A) that actual Type I error probabilities are near the nominal level of the test for a wider range of p_1. The maximum α level that the uncorrected chi-square yields, for the sample sizes used, is .0540 when n = 80 (Table 3).

TABLE 3

MAXIMUM α LEVELS FOR FIVE TESTS AS A FUNCTION OF p_{11}, where $p_{12}=.5-p_{11}$ and $p_{21}=p_{11}$

Sample Size n	Chi-square Uncorrected	Chi-square Corrected	GSK	Likelihood Ratio	Fisher's "Exact"
20	.0512	.0128	.0351	.0824	.0177
40	.0530	.0211	.0433	.0760	.0221
60	.0523	.0278	.0503	.0779	.0282
80	.0540	.0312	.0512	.0767	.0312

In Case II, where none of the cell probabilities exceeds .5, adding the continuity correction again yields very conservative actual α levels (Figure 3B). Although these levels are uniformly higher, over the range of p_1 with an increase in sample size, the maximum α level of the corrected chi-square is only .0312 for a sample of 80.

The GSK test statistic examined in Case II produces surprisingly conservative actual α levels under the first set constraints (Figure 3C). The effect of increasing sample size is generally to increase these levels over the range of p_{11}. The maximum α level achieved, .0512, occurs at n = 80.

Under the same constraints, the likelihood ratio test ensures the least protection from Type 1 errors. Increasing sample size does effectively lower actual α levels; but, if p_{11} or p_{21} is less than .1, some of these levels can exceed .075 (Figure 3D), even with n = 80. For the sample sizes investigated, the likelihood ratio test reached a maximum α level of .0824.

The result of imposing the second set of constraints on the calcuations can be seen in Figures 4A-4D and in Table 4 and can be summarized as follows:

1) maximum α levels for all tests for all sample sizes are lower than under the first set (Table 3),

2) least affected is the uncorrected chi-square,

3) actual α levels of the corrected chi-square and GSK tests generally become more conservative and

4) actual levels still show the likelihood ratio test to be the least effective of the four in controlling Type I errors.

TABLE 4.

MAXIMUM LEVELS FOR FIVE TESTS AS A FUNCTION OF p_{11}, where $p_{12}=.25-p_{11}$ and $p_{21}=3p_{11}$

Sample Size n	Chi-square Uncorrected	Chi-square Corrected	GSK	Likelihood Ratio	Fisher's "Exact"
20	.0482	.0107	.0130	.0800	.0103
40	.0515	.0189	.0303	.0725	.0209
60	.0517	.0242	.0415	.0731	.0246
80	.0511	.0259	.0452	.0710	.0277

Figure 5 illustrates the general agreement between actual α levels of the chi-square test with continuity correction and Fisher's "exact" test defined in the classical sense described by Miettinen [11]. In Figure 5A, we see that, for a total sample of 20 and with constraints: $p_{11} = p_{12}$, $p_{21} = p_{22}$ and $p_{11} + p_{21} = .5$, both tests are conservative with respect to the nominal α level of .05. For p_{11} or $p_{21} <.1$, actual α levels are nearly identical, but, for values of p_{11} near p_{21}, the corrected chi-square is the more conservative test. However, as n increases to 80 (Figure 5B), for the same constraints, there is little difference in the actual α levels of the two tests, although the corrected chi-square is still more conservative over the entire range of p_1.

Under the other set of constraints, i.e. $3p_{11}= p_{12}$, $3p_{21}= p_{22}$, and $p_{11} + p_{21} = .25$, it is evident from Figures 5C-D that, while both tests still produce extremely low actual α levels, the chi-square with continuity

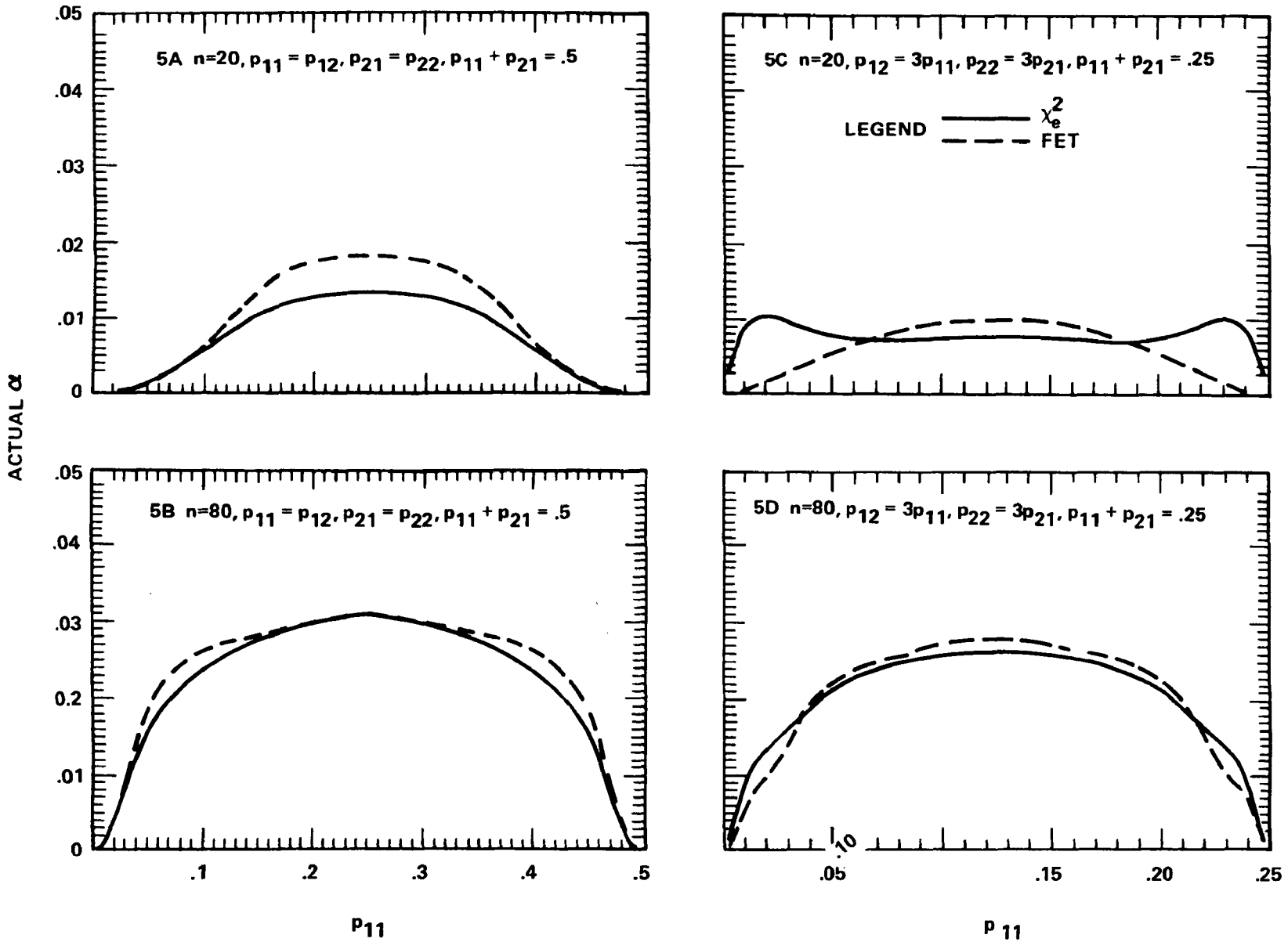

Figure 5. Comparison of α levels of Fisher's "Exact" Test with the corrected Chi-square Test for Case II.

correction is generally more conservative, particularly with p_{11} close to p_{12}. Note that, in Figure 5C, the effect of the continuity correction for p_{11} <.05 or >.2 with a total sample of only 20 is quite pronounced.

As so little is known about the performance of the GSK and likelihood ratio tests in samll samples, we were naturally curious about the power of these tests as well. Because it wasn't feasible to compare the power of these tests under all sets of alternative hypotheses, for each of four sample sizes considered in Case I with $X_{.1} = X_{.2}$, the power was computed by fixing p_2 equal to the proportion for which the actual α levels of the two tests were approximately equal and by letting p_1 vary from 0 to 1.

In Figures 6A-6D, we see that neither of these tests is clearly and uniformly most powerful for any of the sample sizes examined. In fact, with sample sizes of 60 and 80, the two power curves are virtually indistinguishable (Figure 6C-D). For smaller samples, the GSK is more powerful when considering alternate hypotheses where $p_1>p_2$, but the likelihood ratio is more powerful when $p_1<p_2$ (Figure 6A-6B).

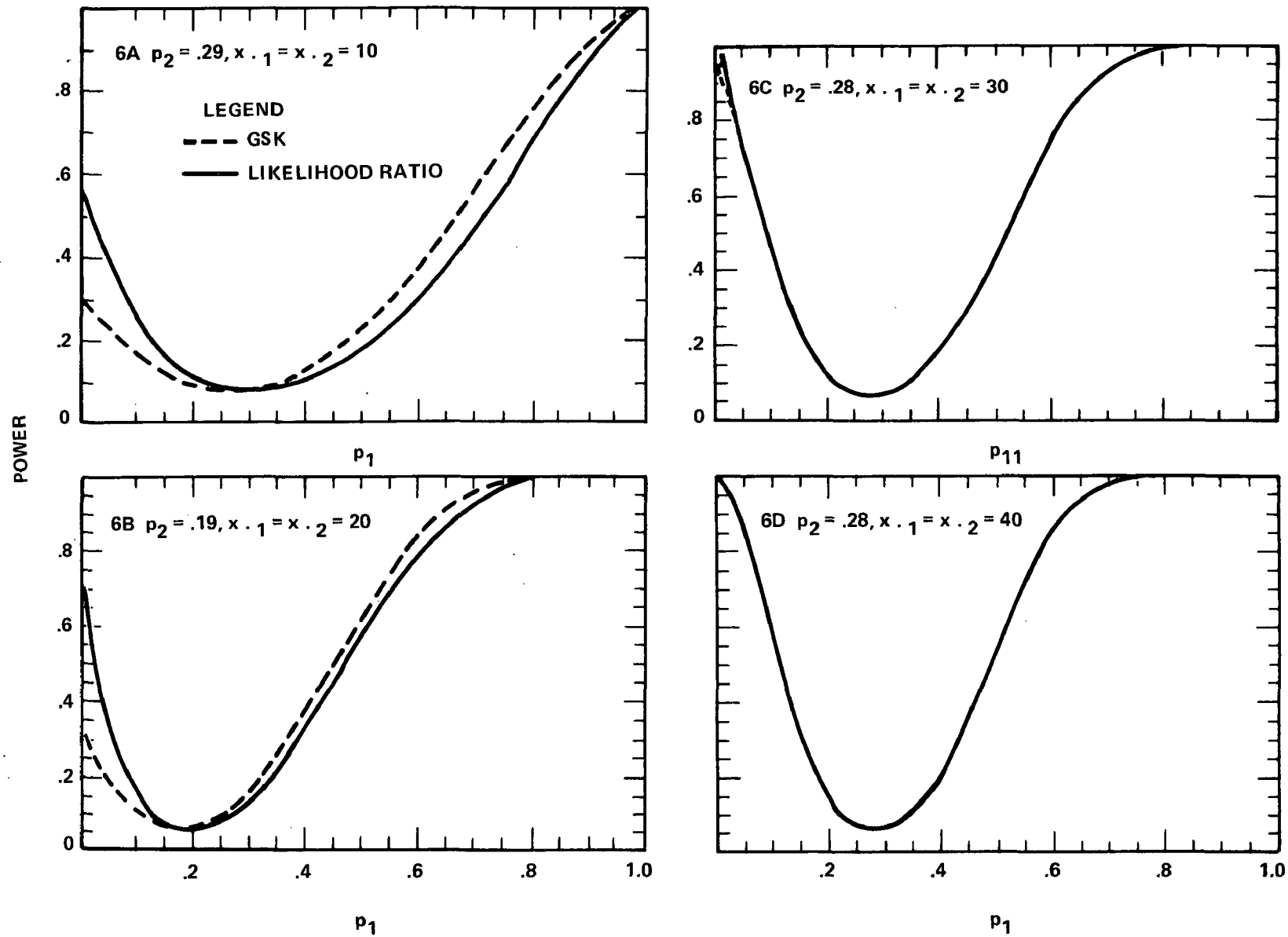

Figure 6. Comparison of the power of the GSK Test vs. the Likelihood Ratio Test. For Case I, where p_2 is chosen so that the α levels of the two tests are the same.

SECTION 4

DISCUSSION

Comparison of the actual α levels of the GSK and likelihood ratio tests indicates that, in general, the GSK test more reliably adheres to a nominal value of .05. For Case I, neither test consistently yields α levels near or below the nominal value. For sample sizes of 30 or 40 from each binomial population, the GSK test appears to provide more acceptable α levels. With unequal sampling, neither test yields levels near the nominal value. For Case II, the GSK test is nearly always conservative while the likelihood ratio test often produces extremely high actual levels. In this case, the GSK test was based on a log-linear model. The log-linear model may be responsible for the more conservative α levels. Although power was compared for so few cases that results may not be generalizable, neither the GSK nor the likelihood ratio test appears to be uniformly more powerful. For larger sample sizes, they were nearly identical.

While this study was undertaken primarily to examine the behavior of the GSK and likelihood ratio tests in small samples, some interesting results occurred in our examination of the other tests. Of all test statistics considered, the uncorrected chi-square was nearly always closest to the nominal value. In fact, of all cases tested, the largest α level that the uncorrected chi-square produced was .059 with actual levels rarely exceeding .055.

Garside and Mack [6] reported exact α levels for the uncorrected chi-square as high as .0607. However, our calculations for the same parameter values produced an α of .0542. Moreover, in every instance where comparisons were possible, our α levels differed substantially from those tabulated by Garside and Mack. Particularly disturbing in these authors' study is the lack of symmetry about p=.5 of these levels when the sample sizes from the two binomial populations differed. Conversely, our results are in complete agreement with those of Eberhardt and Fligner [7] when conditions permitted comparison.

As demonstrated in Figure 5, actual α levels of the corrected chi-square and Fisher's "exact" test are always conservative, and their behavior is quite similar over the entire range of parameters and sample sizes considered. These results are consistent with the observations of others, especially Mantel and Greenhouse [12]. Although only these two tests, of the five studied, provide completely safe α levels, it is very difficult to ignore the performance of the uncorrected chi-square when one is mainly interested in ensuring that the actual level of his test is close to a nominal level of .05.

REFERENCES

[1] House, D. E., J. F. Finklea, C. M. Shy, D. C. Calafiore, W. B. Riggan, J. W. Southwick, L. J. Olsen (1974), Prevalence of Chronic Respiratory Disease Symptoms in Adults: 1970 Survey of Salt Lake Basin Communities. Health Consequences of Sulfur Oxides: A Report from CHESS, 1970-1971. EPA Report No. EPA-65/1-74-4, 2-41 - 2-54. Research Triangle Park, NC, EPA.

[2] Goldberg, H. E., J. F. Finklea, C. J. Nelson, W. B. Steen, R. S. Chapman, D. H. Swanson, A. A. Cohen (1974), Prevalence of Chronic Respiratory Disease Symptoms in Adults: 1970 Survey of New York Communities. Health Consequences of Sulfur Oxides: A Report from CHESS, 1970-1971. EPA Report No. EPA-65/1-74-4, 5-33 - 5-48. Research Triangle Park, NC, EPA.

[3] Finklea, J. F., J. Goldberg, V. Hasselblad, C. M. Shy, C. G. Hayes (1974), Prevalence of Chronic Respiratory Disease Symptoms in Military Recruits: Chicago Induction Center, 1969-1970. Health Consequences of Sulfur Oxides: A Report from CHESS, 1970-1971. EPA Report No. 65/1-74-4, 4-23 - 4-36. Research Triangle Park, NC, EPA.

[4] Grizzle, J. E., C. E. Starmer, G. G. Koch (1969), "Analysis of Categorical Data by Linear Models," Biometrics, 489-504.

[5] Grizzle, J. E. (1967), "Continuity Correction in the X^2 Test for 2 x 2 Tables," The American Statistician, 21, No. 4, 28-32.

[6] Garside, G. R. and C. Mack (1976), "Actual Type 1 Error Probabilities for Various Tests in the Homogeneity Case of the 2 x 2 Table," The American Statistician, 30, No. 1, 18-21.

[7] Eberhardt, K. R. and M. A. Fligner (1977), "A Comparison of Two Tests for Equality of Proportions," The American Statistician, 31, No. 4, 151-155.

[8] Larntz, K. (1978), "Small-Sample Comparisons of Exact Levels for Chi-Squared Goodness-of-Fit Statistics," Journal of the American Statistical Association, 73, No. 362, 253-263.

[9] Goodman, L. (1964), "Simultaneous Confidence Intervals for Contrasts Among Multinomial Populations," Annals of Mathematical Statistics, 35, 716-725.

[10] Kendal, M. G. and A. Stuart (1973), The Advanced Theory of Statistics, 2, 570-572.

[11] Miettinen, O. S. (1974), "Comment on Yates Continuity Correction," JASA, 69, No. 345, 380-382.

[12] Mantel, N. and S. W. Greenhouse (1968), "What is the Continuity Correction?," Amer. Statist., 22, 27-30.

TECHNICAL REPORT DATA
(Please read Instructions on the reverse before completing)

1. REPORT NO. EPA-600/9-80-029	2.	3. RECIPIENT'S ACCESSION NO.
4. TITLE AND SUBTITLE Small Sample Properties of Various Tests on 2 x 2 Tables		5. REPORT DATE May 1980
		6. PERFORMING ORGANIZATION CODE
7. AUTHOR(S) Victor Hasselblad and Andrew G. Stead		8. PERFORMING ORGANIZATION REPORT NO.
9. PERFORMING ORGANIZATION NAME AND ADDRESS Biometry Division, Health Effects Research Laboratory, Research Triangle Park, NC 27711		10. PROGRAM ELEMENT NO. 1AA816
		11. CONTRACT/GRANT NO.
12. SPONSORING AGENCY NAME AND ADDRESS Health Effects Research Laboratory, Office of Research and Development, U.S. Environmental Protection Agency, Research Triangle Park, NC 27711		13. TYPE OF REPORT AND PERIOD COVERED
		14. SPONSORING AGENCY CODE 600/11

15. SUPPLEMENTARY NOTES

16. ABSTRACT

With particular interest in the likelihood ratio tests and tests generated by the categorical analysis methods of Grizzle, Starmer, and Koch, the small sample behavior of actual α levels is graphically illustrated for several asymptotically chi-square statistics when testing is performed on 2 x 2 contingency table data at a nominal level of .05. This behavior is examined for hypotheses testing both homogeneity and independence. In the homogeneity case, the effects on these levels are examined for equal and unequal sample sizes. In the independence case, the effects of different constraints concerning individual cell probabilities are demonstrated. Visual comparisons of the power of the likelihood ratio and the Grizzle, Starmer, Koch related statistics are also provided.

17. KEY WORDS AND DOCUMENT ANALYSIS

a. DESCRIPTORS	b. IDENTIFIERS/OPEN ENDED TERMS	c. COSATI Field/Group
Statistical analysis Chi-square test	Categorical analysis methods Likelihood ratio tests Small sample behavior	12A

18. DISTRIBUTION STATEMENT RELEASE TO PUBLIC	19. SECURITY CLASS *(This Report)* UNCLASSIFIED	21. NO. OF PAGES 16
	20. SECURITY CLASS *(This page)* UNCLASSIFIED	22. PRICE

EPA Form 2220-1 (Rev. 4-77) PREVIOUS EDITION IS OBSOLETE