

ON TESTING HYPOTHESES CONCERNING STANDARDIZED
MORTALITY RATIOS AND/OR INDIRECT ADJUSTED RATES*

By

Lawrence L. Kupper
and
David G. Kleinbaum

Department of Biostatistics
University of North Carolina at Chapel Hill
Institute of Statistics Mimeo Series No. 709

September 1970

ON TESTING HYPOTHESES CONCERNING STANDARDIZED
MORTALITY RATIOS AND/OR INDIRECT ADJUSTED RATES*

Lawrence L. Kupper and David G. Kleinbaum

University of North Carolina at Chapel Hill

1. Introduction.

Suppose that there are p test populations $\pi_1, \pi_2, \dots, \pi_p$, a standard population π_0 , and c categories, where the term "test population" denotes a collection of individuals under study. We shall regard a test population as a sample from some universe, for only then is there any point in discussing statistical estimation and hypothesis testing (see Section I of [2] for comments along this line).

Let

n_{ij} = the number of individuals in category j of π_i ,

p_{ij} = the true probability of death in category j of π_i ,

d_{ij} = the observed number of deaths in category j of π_i ,

$i=0,1,\dots,p$, $j=1,2,\dots,c$. Then, the standardized mortality ratio $SMR(i)$ for population π_i is defined to be

*Work supported by National Institutes of Health Grant No. GM-12868.

$$(1.1) \quad \text{SMR}(i) = \frac{\sum_{j=1}^c n_{ij} p_{ij}}{\sum_{j=1}^c n_{ij} p_{oj}}, \quad i=0,1,\dots,p.$$

(Note that $\text{SMR}(0) = 1$.) In this paper we shall be concerned with testing the following general hypothesis, namely

$$(1.2) \quad H_0: \text{SMR}(i_1) = \text{SMR}(i_2) = \dots = \text{SMR}(i_k),$$

where $1 \leq i_1 < i_2 < \dots < i_k \leq p$ and $2 \leq k \leq p$.

The indirect adjusted death rate $P(i)$ for π_i is

$$P(i) = \sum_{j=1}^c \omega_{ij} p_{ij}, \quad i=0,1,\dots,p,$$

where

$$\omega_{ij} = \left(\frac{n_{ij}}{\sum_{j=1}^c n_{ij} p_{oj}} \right) \left(\frac{\sum_{j=1}^c n_{oj} p_{oj}}{\sum_{j=1}^c n_{oj}} \right).$$

Since $P(i) = \alpha_0 \text{SMR}(i)$, where $\alpha_0 = \frac{\sum_{j=1}^c n_{oj} p_{oj}}{\sum_{j=1}^c n_{oj}}$ is independent of i , it is clear that testing the hypothesis H_0 given by (1.2) is equivalent to testing the hypothesis

$$H'_0: P(i_1) = P(i_2) = \dots = P(i_k).$$

Hence, in what follows we can, without loss of generality, restrict ourselves to considering tests of H'_0 .

Finally, although we have presented the problem in a mortality-type setting, the techniques developed below can validly be used in any situation

where the response is dichotomous (e.g., was or was not a migrant, did or did not have cancer).

2. Choice of π_0 and its effect on tests of H_0 .

When the standard population π_0 is chosen independently of the test populations $\pi_1, \pi_2, \dots, \pi_p$, then it is reasonable to assume, for all practical purposes, that the quantities $p_{0j} = d_{0j}/n_{0j}$ are non-stochastic (c.f., [2], [4], [5]). An example of such a situation would be one in which the test populations are all the different counties in England in 1963 and the standard is the U. S. population in 1963 or the 1945 population of England. In this case the best estimate of $SMR(i)$ is

$$(2.1) \quad \hat{SMR}(i) = \frac{\sum_{j=1}^c n_{ij} \hat{p}_{ij}}{\sum_{j=1}^c n_{ij} p_{0j}}, \quad i=1, 2, \dots, p,$$

where $\hat{p}_{ij} = d_{ij}/n_{ij}$ is approximately normally distributed for large n_{ij} , has mean p_{ij} and variance $p_{ij}(1-p_{ij})/n_{ij}$, and is uncorrelated with $\hat{p}_{i'j'}$, unless both $i=i'$ and $j=j'$. The corresponding best estimate of $P(i)$ is $\hat{P}(i) = \sum_{j=1}^c \omega_{ij} \hat{p}_{ij}$.

However, the situation is quite different when the standard population is formed by pooling the test populations together so that p_{0j} has the structure

$$(2.2) \quad p_{0j} = \frac{\sum_{i=1}^p n_{ij} p_{ij}}{\sum_{i=1}^p n_{ij}}, \quad j=1, 2, \dots, c.$$

In this case not only the numerator but also the denominator of $SMR(i)$ must be estimated using the \hat{p}_{ij} 's obtained from the test populations; in particular, the best estimate of $SMR(i)$ is then given by

$$(2.3) \quad \widehat{SMR}(i) = \sum_{j=1}^c n_{ij} \widehat{p}_{ij} / \sum_{j=1}^c n_{ij} \widehat{p}_{oj}, \quad i=1,2,\dots,p,$$

where $\widehat{p}_{oj} = \sum_{i=1}^p n_{ij} \widehat{p}_{ij} / \sum_{i=1}^p n_{ij}$. The corresponding best estimate of $P(i)$ is given by $\widehat{P}(i) = \sum_{j=1}^c \widehat{\omega}_{ij} \widehat{p}_{ij}$, where $\widehat{\omega}_{ij}$ is obtained by substituting \widehat{p}_{oj} for p_{oj} and $n_{.j} = \sum_{i=1}^p n_{ij}$ for n_{oj} in the expression for ω_{ij} .

The use of a "pooled standard" of the form (2.2) is not without precedent. For example, researchers associated with the Evans County Coronary Heart Disease Study [1] have compared the SMR of white males having low systolic blood pressure with that of white males having high systolic blood pressure, the categories being age groups and the standard population consisting of all white males in Evans County, Georgia. Similar pooled standards were used to study the effects of other physiological factors as well. In general, the use of such a standard is dictated when the response under study (e.g., the incidence of coronary heart disease) has not been the subject of a previous investigation conducted under comparable conditions.

It is important to emphasize the distinction between the above two choices of the standard population because the different estimates (2.1) and (2.3) necessitate the use of different procedures for testing H_0 . These approaches are described in the next two sections.

3. Testing H_0 when π is chosen independently of the test populations.

If we define the vector

$$\widehat{\theta}' = (\widehat{SMR}(1), \widehat{SMR}(2), \dots, \widehat{SMR}(p)),$$

where $\widehat{SMR}(i)$ is given by (2.1), then it follows that

$$E(\widehat{\theta}) = \theta \quad \text{and} \quad \text{Var}(\widehat{\theta}) = \text{diag}(\sigma_1^2, \sigma_2^2, \dots, \sigma_p^2) = D,$$

where $\underline{\theta}' = (\text{SMR}(1), \text{SMR}(2), \dots, \text{SMR}(p))$ and

$$\sigma_i^2 = \sum_{j=1}^c n_{ij} p_{ij} (1-p_{ij}) / (\sum_{j=1}^c n_{ij} p_{oj})^2, \quad i=1,2,\dots,p.$$

In this framework, the hypothesis H_0 can equivalently be written as

$$H_0: C \underline{\theta} = \underline{0},$$

where the $(k-1) \times p$ matrix C has as its $(j-1)$ -th row the vector

$\underline{c}'_{j-1} = (0, \dots, 0, 1, 0, \dots, 0, -1, 0, \dots, 0)$ having +1 in the i_1 -th position, -1 in the i_j -th position, and zeros elsewhere, $j=2,3,\dots,k$. From the fact that

$$T_1 = \hat{\theta}' C' (\hat{C} D \hat{C}')^{-1} C \hat{\theta}$$

is approximately distributed as a central χ^2 variable with $(k-1)$ degrees of freedom when H_0 is true, where $\hat{D} = \text{diag}(\hat{\sigma}_1^2, \hat{\sigma}_2^2, \dots, \hat{\sigma}_p^2)$ is the matrix obtained from D by replacing p_{ij} by \hat{p}_{ij} everywhere it appears, we would reject H_0 at the α level of significance if $T_1 \geq \chi_{k-1; 1-\alpha}^2$, where $\chi_{k-1; 1-\alpha}^2$ is the upper 100(1- α)% point of the central χ_{k-1}^2 distribution.

For the special case of H_0 when $k=p$ (so that we are testing the equality of all p SMR's or all p indirect rates), C is simply the $(p-1) \times p$ matrix

$$(3.1) \quad \begin{bmatrix} 1 & -1 & 0 & 0 & \dots & 0 & 0 \\ 1 & 0 & -1 & 0 & \dots & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 1 & 0 & 0 & 0 & \dots & 0 & -1 \end{bmatrix}.$$

When $k=2$, we may alternatively use as the test statistic

$$Z = [\widehat{SMR}(i_1) - \widehat{SMR}(i_2)] / (\hat{\sigma}_{i_1}^2 + \hat{\sigma}_{i_2}^2)^{1/2},$$

which is approximately distributed as a standard normal random variable when H_0 is true. In this case we would reject H_0 at the α level of significance when $|Z| \geq Z_{1-\alpha/2}$, where $Z_{1-\alpha/2}$ is the 100(1- $\alpha/2$)% point of the standard normal distribution.

The use of T_1 to test hypotheses which involve an $SMR(i)$ for which \hat{p}_{ij} is either 0 or 1 for every j requires special attention since the estimated variance $\hat{\sigma}_i^2$ of $\widehat{SMR}(i)$ is zero. Such a situation can be handled by either choosing the estimate of p_{ij} according to some criterion other than maximum likelihood (e.g., $(d_{ij}+1)/(n_{ij}+2)$ is the Bayes estimate for a uniform prior on p_{ij}) or by deciding not to test hypotheses about $SMR(i)$.

4. Testing H_0 when π is formed from the test populations.

When p_{oj} is of the form (2.2) so that the estimate of $SMR(i)$ is given by (2.3), the procedure for testing H_0 depends on the relationships expressed in the following lemmas.

Lemma 4.1. When p_{oj} has the form (2.2), then $\sum_{i=1}^P a_i SMR(i) = 1$, where $a_i = \sum_{j=1}^C n_{ij} p_{oj} / \sum_{i=1}^P \sum_{j=1}^C n_{ij} p_{oj}$.

Proof. From (1.1), $\sum_{i=1}^P a_i SMR(i) = \sum_{i=1}^P (\sum_{j=1}^C n_{ij} p_{ij} / \sum_{i=1}^P \sum_{j=1}^C n_{ij} p_{oj})$
 $= \sum_{j=1}^C (\sum_{i=1}^P n_{ij} p_{ij}) / \sum_{j=1}^C \sum_{i=1}^P n_{ij} p_{oj}$, and the result then follows by using (2.2).

Lemma 4.2. When \hat{p}_{oj} has the form (2.2), then $SMR(1)=SMR(2)=\dots=SMR(p)$ if and only if there is a subset of $(p-1)$ SMR's all equal to one.

Proof. If $SMR(1)=SMR(2)=\dots=SMR(p)=\mu$, say, then, from Lemma 4.1, $\mu \sum_{i=1}^p a_i = 1 = \mu$ since $\sum_{i=1}^p a_i = 1$. If there is a subset of $(p-1)$ SMR's all equal to one, then, taking this subset to be the first $(p-1)$ SMR's, it follows from Lemma 4.1 that $\sum_{i=1}^{p-1} a_i + a_p SMR(p) = 1$, or $SMR(p) = (1 - \sum_{i=1}^{p-1} a_i) / a_p = a_p / a_p = 1$. This completes the proof.

Lemma 4.3. Let $W_{ii'} = \sum_{j=1}^c n_{ij} n_{i'j} (p_{ij} - p_{i'j}) / n_{.j}$ and define $W_i = \sum_{i'=1}^p W_{ii'}$, $i, i'=1, 2, \dots, p$.

Then,

- (i) $W_{ii'} = -W_{i'i}$ and $W_{ii'} = 0$ if $i=i'$,
- (ii) $\sum_{i=1}^p W_i = 0$,
- (iii) when p_{oj} has the form (2.2), $SMR(i)=1$ if and only if $W_i = 0$, $i=1, 2, \dots, p$.

Proof. (i) Trivial.

$$(ii) \sum_{i=1}^p W_i = \sum_{i=1}^p \sum_{j=1}^c n_{ij} p_{ij} - \sum_{i'=1}^p \sum_{j=1}^c n_{i'j} p_{i'j} = 0.$$

(iii) Using (1.1) and (2.2), we have

$$\begin{aligned}
 \text{SMR}(i)-1 &= \left(\sum_{j=1}^c n_{ij} p_{ij} / \sum_{j=1}^c n_{ij} p_{oj} \right) - 1 \\
 &= \left[\sum_{j=1}^c n_{ij} p_{ij} - \sum_{j=1}^c n_{ij} \left(\sum_{i'=1}^p n_{i'j} p_{i'j} / \sum_{i'=1}^p n_{i'j} \right) \right] / \sum_{j=1}^c n_{ij} p_{oj} \\
 &= \left[\sum_{j=1}^c n_{ij} \left(\sum_{i'=1}^p n_{i'j} p_{ij} - \sum_{i'=1}^p n_{i'j} p_{i'j} \right) / n_{.j} \right] / \sum_{j=1}^c n_{ij} p_{oj} \\
 &= \sum_{i'=1}^p \left[\sum_{j=1}^c n_{ij} n_{i'j} (p_{ij} - p_{i'j}) / n_{.j} \right] / \sum_{j=1}^c n_{ij} p_{oj} \\
 &= \sum_{i'=1}^p W_{ii'} / \sum_{j=1}^c n_{ij} p_{oj} = W_i / \sum_{j=1}^c n_{ij} p_{oj}. \quad \text{This completes the proof.}
 \end{aligned}$$

So, when p_{oj} is of the form (2.2), it follows from Lemma 4.2 and from part (iii) of Lemma 4.3 that testing the hypothesis that $\text{SMR}(1)=\text{SMR}(2)=\dots=\text{SMR}(p)$ ($=1$) or that $P(1)=P(2)=\dots=P(p)(=\alpha_0)$ is equivalent to testing the hypothesis that $W_i=0$ for every i , $i=1,2,\dots,p$. The appropriate test statistic for this latter hypothesis is

$$T_2 = \hat{W}' C' (C \hat{\Sigma} C')^{-1} C \hat{W},$$

where \hat{W}' and $\hat{\Sigma}$ are the estimates of $W' = (W_1, W_2, \dots, W_p)$ and $\Sigma = ((\text{cov}(\hat{W}_i, \hat{W}_{i'})))_{i,i'=1}^p$ obtained by putting \hat{p}_{ij} for p_{ij} , and where C is given by (3.1). (From part (ii) of Lemma 4.3, it follows that one can alternatively use for C the matrix obtained by deleting any one row of the identity matrix of order p .) The statistic T_2 is approximately distributed as a central χ^2 variate with $(p-1)$ degrees of freedom when the hypothesis that $W=0$ is true.

The determination of the elements of $\hat{\Sigma}$ is somewhat tedious but straightforward. In particular,

$$\text{cov}(\hat{W}_i, \hat{W}_{i'}) = \sum_{\ell=1}^p \sum_{\ell'=1}^p \text{cov}(\hat{W}_{i\ell}, \hat{W}_{i'\ell'})$$

can be evaluated directly from the following formulae (see part (i) of Lemma 4.3):

$\text{cov}(\hat{W}_{i\ell}, \hat{W}_{i'\ell'}) = 0$ if either $i=\ell$, $i'=\ell'$, or both, or if i, ℓ, i' and ℓ' are all different;

$\text{cov}(\hat{W}_{i\ell}, \hat{W}_{i'\ell'}) = -\text{cov}(\hat{W}_{i\ell}, \hat{W}_{\ell'i'}) = -\text{cov}(\hat{W}_{\ell i}, \hat{W}_{i'\ell'}) = \text{cov}(\hat{W}_{\ell i}, \hat{W}_{\ell'i'})$;

$\text{cov}(\hat{W}_{i\ell}, \hat{W}_{i\ell'}) = \sum_{j=1}^c n_{ij} n_{\ell j} n_{\ell' j} p_{ij} (1-p_{ij}) / n_{\cdot j}^2$ if i, ℓ and ℓ' are all different; and, for $i \neq \ell$,

$$\text{var}(\hat{W}_{i\ell}) = \sum_{j=1}^c \frac{n_{ij}^2 n_{\ell j}^2}{n_{\cdot j}^2} \left[\frac{p_{ij} (1-p_{ij})}{n_{ij}} + \frac{p_{\ell j} (1-p_{\ell j})}{n_{\ell j}} \right].$$

For the special case $p=2$, the test statistic takes the form $\hat{W}_{12}^2 / \text{var}(\hat{W}_{12})$, which is approximately distributed as χ_1^2 when $\text{SMR}(1)=\text{SMR}(2)$. This case has been studied by Cochran [3] and Quade (unpublished).

Unfortunately, the above procedure cannot be used to test the general hypothesis H_0 given by (1.2) when k is to be strictly less than p . To illustrate the difficulty, consider the case where $k=2 < p$ and we wish to test $H_0: \text{SMR}(i_1) = \text{SMR}(i_2)$. Then, because p exceeds k , we cannot claim that this hypothesis is equivalent to the hypothesis that $W_{i_1} = W_{i_2} = 0$, which is necessary for the test statistic T_2 to be appropriate. A reasonable solution to this problem is suggested by the findings presented in the next section.

5. Some experimental results.

Computer programs have been prepared which calculate SMR's and indirect rates for $p \leq 10$ populations and $c \leq 20$ categories. One of these programs is written in Fortran IV for the IBM 360 computer at the Triangle University

Computation Center (TUCC) and the other in Time-Sharing Fortran for the Call-A-Computer system. These programs allow the user to specify the standard population or to instruct the computer to form the standard by pooling the test populations. The user of either program may test hypotheses of the form (1.2) as described in Sections 3 and 4.

These programs were used to compare the two test procedures presented in this paper in the situation where the standard population is formed by pooling the test populations. The results obtained should be of considerable interest to those researchers who have incorrectly used the test procedure described in Section 3 when they should have employed that of Section 4, for it is conceivable that inferences drawn from using the "incorrect" test statistic could be reversed when the "correct" one is used.

The data used in this computer study was obtained from the Evans County Coronary Heart Disease Study mentioned in [1]. There were 38 different data sets, each partitioned into the same $c=7$ age groups; 25 of the sets were based on two populations, 7 on three and 6 on four. The hypothesis (1.2) for $k=p$ was tested and the values of the test statistics T_1 and T_2 for each data set are presented in Table I. None of the data sets gave estimated SMR's of zero (see the last paragraph of Section 3).

It can be seen from Table I that, despite the rationale for preferring the test procedure of Section 4, both test statistics give almost exactly the same numerical values in each case. This leads us to feel that a test of the hypothesis (1.2) when the standard is pooled from the test populations and the "incorrect" test procedure of Section 3 is used will yield fairly reliable results not only for the case $k=p$ but also when $k < p$.

6. ~~Summary~~.

This paper describes tests of hypotheses of the form $H_0: \text{SMR}(i_1)=\text{SMR}(i_2)=\dots=\text{SMR}(i_k)$ or $H'_0: P(i_1)=P(i_2)=\dots=P(i_k)$, where $1 \leq i_1 < i_2 < \dots < i_k \leq p$ and $2 \leq k \leq p$, when there are p test populations and c categories for each population. The distinction is made between a standard population which is chosen independently of the test populations and one which is formed by pooling the test populations. Different test procedures are appropriate for these two situations. When the pooled standard is used, the appropriate test procedure is applicable only when $k=p$. Experimental evidence is given showing that when the pooled standard is used both the "correct" and "incorrect" test procedures lead to the same conclusion concerning H_0 (or H'_0) for the case $k=p$. The recommendation is made, therefore, to use the "incorrect" test procedure when the standard is formed by pooling the test populations even when $k < p$.

TABLE I

VALUES OF T_1 AND T_2 OBTAINED BY TESTING H_0
FOR $k=p$ WHEN THE STANDARD IS FORMED BY
POOLING THE TEST POPULATIONS

<u>NUMBER OF POPULATIONS p</u>	<u>"INCORRECT" T_1</u>	<u>"CORRECT" T_2</u>
2	28.2136	28.7157
2	16.9346	17.4983
2	21.2930	21.6004
2	10.2976	10.1884
2	10.4231	10.8275
2	1.7373	1.7946
2	3.1895	3.2292
2	7.6659	7.3038
2	0.1513	0.1533
2	0.0475	0.0476
2	1.2320	1.2488
2	0.1386	0.1450
2	1.4684	1.4437
2	1.9818	1.9965
2	0.1872	0.1947
2	1.3005	1.2828
2	1.4704	1.4462
2	0.2089	0.2131
2	0.4434	0.4296
2	0.5522	0.5302
2	2.2685	2.1455
2	0.0042	0.0043
2	0.0361	0.0466
2	2.6843	2.7078
2	0.0016	0.0016
3	3.5495	3.5685
3	8.1122	7.8162
3	10.4558	10.8670
3	2.5145	2.4496
3	10.6485	10.5651
3	17.0558	17.6127
3	21.6299	21.9669
4	23.9671	24.6090
4	20.3859	20.9092
4	16.6307	16.4557
4	7.1490	7.0270
4	9.7703	9.6688
4	3.6639	3.6656

REFERENCES

- [1] CASSEL, J. C., HEYDEN, S., TYROLER, H. A., CORNONI J., BARTEL, A. and HAMES, C. G. (1970). Differences in probability of death related to physiological parameters in a biracial total community study (Evans County, Georgia). To be published.
- [2] CHIANG, C. L. (1961). Standard error of the age-adjusted death rate. U. S. Department of Health, Education and Welfare, Vital Statistics - Special Reports, 47, 271-285.
- [3] COCHRAN, W. G. (1954). Some methods of strengthening the common χ^2 tests. Biometrics, 10, 417-451.
- [4] HILL, A. B. (1956). Principles of Medical Statistics. New York, Oxford University Press.
- [5] KEYFITZ, N. (1966). Sampling variance of standardized mortality rates. Human Biology, 38, 309-317.