

Statistical Reanalysis of Jewish Priests' and Non-Priests' Haplotypes Using Exact Methods

SAGE Open
1–3
© The Author(s) 2011
DOI: 10.1177/2158244011413475
<http://sgo.sagepub.com>


Shlomo S. Sawilowsky¹

Abstract

Researchers in an article appearing in *Nature* used asymptotic (i.e., large sample) chi-square tests in analyzing haplotypes of Y chromosomes using the polymerase chain reaction applied to genomic DNA from male Israeli, North American, and British Jews. The use of classical methods for analyzing extremely sparse contingency tables is frequently done, but with the advent of statistical software capable of conducting exact tests, researchers should certainly cease relying on outdated methods for small sample analyses. A reanalysis was conducted using modern statistical methods. Results and implications for using exact tests are discussed.

Keywords

anthropology, computer methods, data processing and interpretation, religious studies, research methodology and design, research methods, statistical theory and tests

Introduction

Skorecki et al. (1997) reported on neutral DNA markers (mitochondrial and Y-chromosome) for male Jews who were *Cohanim* (s., “*Cohain*”; a Hebrew term translated as Priests, referring to descendants of the subtribe of Levi restricted to the lineage of Aaron, the brother of Moses). According to Jewish law, a strict patrilineal Priestly descent is required to perform sacramental duties. Skorecki et al. found support for the presence of these markers, but the statistical techniques they used were not the best available. More precise methods are used here to strengthen the conclusions of the earlier study: There is a genetic marker for the male descendents of Aaron in Jews today who are identified as *Cohanim*.

The haplotypes of “Y chromosomes using the polymerase chain reaction applied to genomic DNA isolated from buccal swab samples from Israeli, North American, and British Jews” (Skorecki et al., 1997, p. 32), were examined. A total of 188 study participants were further identified as being of Ashkenazic (Western [e.g., German] or Central [e.g., Russian] European) or Sephardic (e.g., Iberian, North African, and Yemenite) descent. Because they only reported percentages, for clarity the frequencies and totals were computed and are presented in Table 1.

Data analysis by Skorecki et al. (1997) consisted of the classical asymptotic χ^2 test of goodness of fit. It was employed on the proportion of Y-chromosome haplotypes for Priests versus non-Priests. The obtained p value was reported as $<.001$, distinguishing the Priestly class from the non-Priests. Further analyses showed this difference was apparent in both

the Ashkenazic ($p < .01$) and Sephardic ($p < .01$) community subsamples.

However, Hirji (2006), Moses, Emerson, and Hosseini (1984), among many others, noted that exact tests are preferable for small data sets such as those frequently obtained in medical research. The reason was underscored by Mehta and Patel (1995), who noted that when contingency tables are sparse, “the usual chi-squared asymptotic distribution . . . is not likely to yield accurate p -values” (p. 577). This problem is exacerbated when more than 20% of the cells have expected frequencies less than five (Cochran, 1954; Everitt, 2000; Siegal & Castellan, 1988). Although the existence of exact tests was available via commercial software when the Skorecki et al. (1997) study was published, even today these methods are rarely invoked.

Purpose of the Study

The purpose of this study is to reanalyze the data in the Skorecki et al. (1997) experiment. The intent is to demonstrate the utility of using modern, computer-intensive exact methods over traditional asymptotic tests.

¹Wayne State University, Detroit, MI, USA

Corresponding Author:

Shlomo S. Sawilowsky, Evaluation & Research, Wayne State University,
371 EDUC, Detroit, MI 48202, USA
Email: shlomo@wayne.edu

Table 1. Haplotype Frequencies and Totals for Jewish Priests and Non-Priests for All, Ashkenazic, and Sephardic Communities

Alleles	All		Ashkenazic		Sephardic	
	Priests (n = 68)	Non-priests (n = 120)	Priests (n = 44)	Non-Priests (n = 81)	Priests (n = 24)	Non-priests (n = 39)
YAP ⁻ DYS19A	11	11	9	6	2	5
B	37	39	20	26	17	13
C	11	36	10	22	1	14
D	6	10	4	9	2	1
E	2	2	0	2	2	0
YAP ⁺ DYS19	1	22	1	16	0	6
Total	N = 188		n = 125		n = 63	

Results

A reanalysis of the data provided by Skorecki et al. (1997) on the frequencies computed in Table 1 above replicated their significant asymptotic results ($\chi^2 = 20.83$, $df = 5$, $p = .0006$) for Priests versus non-Priests. Exact tests (performed with *StatXact*, 2010) also confirmed this finding (χ^2 exact $p = .0006$) and separately for the Sephardic Jews ($\chi^2 = 18.92$, $df = 5$, exact $p = .0005$).

However, divergent results ($p > .01$) were obtained with exact methods for Ashkenazic Jews, where the hypothesis of no difference between Priests and non-Priests was retained at the nominal $\alpha = .01$ level ($\chi^2 = 13.25$, $df = 5$, exact $p = .0164$). Interestingly, the asymptotic χ^2 results were not replicated ($\chi^2 = 13.25$, $df = 5$, $p = .0211$).

The mixed findings are likely the result of the low power associated with the χ^2 test. A more powerful approach is to conduct an exact $2 \times c$ analysis of the six haplotypes for Priests versus non-Priests, stratified based on Ashkenazic versus Sephardic descent. The result of this test was significant (Permutation test exact $p = 0.0001$) and offers unequivocal statistical evidence based on the study sample of the homogeneity within the priestly lineage in both Jewish subcommunities.

Skorecki et al. (1997) did not assess the differences between Ashkenazic and Sephardic Jews, either by combining or controlling for Priestly versus non-Priestly status. Therefore, that analysis was also computed using exact methods. Combining categories was not significant ($\chi^2 = 3.241$, $df = 5$, exact $p = .6823$). Similarly, controlling for the distinction with a $2 \times c$ analysis of alleles for Ashkenazim versus Sephardim, stratified based on Priest versus non-Priest, was not significant (Permutation test exact $p = .1771$).

Discussion

The purpose of this article was to caution against relying on asymptotic theory when exact methods are available. The reanalysis using modern exact methods presents a striking

demonstration of the solidarity between the two major Jewish subcultures in preserving the strict patrilineal requirement for the Priestly status. This is remarkable considering the Jewish exile began in 574 B. C. E. when the first $2\frac{1}{2}$ tribes (i.e., Gad, Reuben, and half of Menasheh) and the Priests who lived among them were expelled from Israel (II Kings 15:29). This finding was overlooked simply by relying solely on classical statistical methods.

Although this reanalysis is a reminder to researchers to eschew classical asymptotic methods in favor of modern methods, a caution is nevertheless in order. Researchers should not assume (as software vendors frequently claim) that the advantage of exact methods is that they will necessarily lead to *smaller* p values and a greater likelihood to reject the null hypothesis. In fact, as was the case in one of the reanalyses above, due to estimation procedures (i.e., interpolation) used in creating the classic asymptotic tabled values commonly found in research and statistic textbooks and generic statistical software, they are as likely to lead to a smaller p value as are exact methods. However, the advantage favoring exact methods is that they supply the *correct* p value and, hence, lead to more accurate interpretation of statistical results.

Declaration of Conflicting Interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) received no financial support for the research and/or authorship of this article.

References

- Cochran, W. G. (1954). Some methods for strengthening the common χ^2 tests. *Biometrics*, 10, 417-454.
- Everitt, B. S. (2000). *The analysis of contingency tables* (2nd ed.). Boca Raton, FL: Chapman & Hall/CRC.
- Hirji, K. F. (2006). *Exact analysis of discrete data*. Boca Raton, FL: Chapman & Hall/CRC.

- Mehta, C., & Patel, N. (1995). *StatXact 3.0 for Windows: Statistical software for exact nonparametric inference. User manual*. Cambridge, MA: Cytel Software.
- Moses, L. E., Emerson, J. D., & Hosseini, H. (1984). Statistics in practice: Analyzing data from ordered categories. *New England Journal of Medicine*, 311, 442-448.
- Siegel, S., & Castellan, N. J. (1988). *Nonparametric statistics for the behavioral sciences* (2nd ed.). New York, NY: McGraw-Hill.
- Skorecki, K., Selig, S., Blazer, S., Bradman, R., Bradman, N., Waburton, P. J., . . . Hammer, M. F. (1997). Y chromosomes of Jewish priests. *Nature*, 385, 32.
- StatXact*. (2010). Version 9. Cambridge, MA: Cytel.

Bio

Shlomo S. Sawilowsky is a WSU Distinguished Faculty Fellow, author of *Real Data Analysis* (2006, ISBN 978-1-59311-565-4), and founding editor of the *Journal of Modern Applied Statistical Methods* (www.jmasm.org).