JBL 103/1 (1984) 11–21

BIBLICAL TEXTS AND STATISTICAL ANALYSIS: ZECHARIAH AND BEYOND

STEPHEN L. PORTNOY
University of Illinois, Urbana, IL 61801
DAVID L. PETERSEN
Iliff School of Theology, Denver, CO 80201

I. Introduction

It is an understatement to maintain that pluralism is the hallmark of contemporary biblical studies. Rarely have there been so many ways to study biblical texts. Lower and higher criticisms have been supplemented in recent decades with, among others, structuralism, canonical criticism, literary criticism, social scientific approaches, and statistical methods of analysis. Despite, and perhaps because of, this wealth of critical approaches, a problem has arisen. Though practitioners of the hitherto standard crafts, e.g., textual or form criticism, are often tolerant of their more avant garde colleagues, they do not regularly engage these methods in dialogue, nor do they regularly assess the validity of claims made by those using newer approaches. This situation seems to be especially true for those methods that use a specialized vocabulary—structuralism with its actantial models or statistical analysis with its chi-square tests. Conversely, advocates of the more recently articulated interpretive styles sometimes ignore or disparage the results of earlier interpretive modes. The result is that the various methods of biblical study are isolated from one another and each operates in a vacuum. Only when the results of one method appear to challenge directly the results of another method does dialogue appear necessary, and even then it does not always ensue.

The authors of this study—one a statistician, the other a biblical scholar—wish to accept the burden of engaging in such dialogue between and about methods, at least in part because of recent journalistic reports concerning statistical analysis of Genesis. Yehuda Radday, one of the investigators in that project, has been quoted as follows: "It is most probable that the book of Genesis was written by one person."[1] This assertion challenges some of the most widely accepted conclusions of the

[1] So *Newsweek* (28 September 1981) 59. See also an article in the *New York Times* (8 November 1981) 10.

higher critical study of biblical literature, and it seems to demand dialogue not only about the character of the statistical procedures used in the project but also about the place of statistical analysis within the full panoply of methods available for the study of the Bible.

The most concise presentation of Radday's statistical methods is in an article that focuses on the unity of authorship in the book of Zechariah.[2] It is upon this articulation of statistical analysis that the authors of this paper have concentrated. After carefully assessing Radday and Wickmann's study, we make the following three observations:

1. Radday and Wickmann (hereafter R–W) have made some major statistical errors that invalidate any firm conclusions their study on Zechariah may proffer.

2. After the procedures for examining statistically the issue of authorship are assessed, it is possible to propose statistical models that are appropriate to the task of analyzing the book of Zechariah. And despite the fact that the book of Zechariah almost surely represents too small and too heterogeneous a body of literature for statistical methods alone to provide definitive conclusions concerning unity of authorship, it appears, on statistical grounds, that Zechariah 1–8 may be distinguished from 9–11 and from 12–14, and that Zechariah 9–11 and 12–14 may also be distinguished from each other.

3. Use of other methods, i.e., tradition history and form criticism in the case of Zechariah, should be integrated into a discussion of the book's unity. Proper pluralism in biblical study entails dialogue between the various methodological perspectives and a willingness to engage in mutual self-correction. Satisfactory conclusions about the complexities of texts from the Iron Age or the Persian period rarely follow from one method pursued in heroic isolation from other perspectives.[3]

[2] Yehuda T. Radday and Dieter Wickmann, "The Unity of Zechariah examined in the Light of Statistical Linguistics," *ZAW* 87 (1975) 30–55. A more recent paper by these same authors ("Vocabulary Richness in Post-Exilic Prophetic Books," *ZAW* 92 [1980] 333–46) introduces a different measure of linguistic style, the "word frequency profile." Using this measure, they draw conclusions similar to those contained in this study.

[3] So, for example, the text of Jeremiah presents difficult problems that require a sophisticated combination of text-, literary-, and redaction-critical insights. See E. Tov, "L'incidence de la critique textuelle sur la critique littéraire dans le livre de Jérémie," *RB* 79 (1972) 189–99. Or, in a quite different sort of example, Robert Alter integrates insights drawn from higher critical study of the Bible into his avowedly literary-critical work on biblical prose, *The Art of Biblical Narrative* (New York: Basic Books, 1981). To be sure, one perspective on a given text can yield valuable insights. And yet, by itself, such an exercise comprises only one part of the complete interpretive task.

II. The Statistical Errors

R–W begin their analysis by introducing thirty-six measures or criteria of literary style. For example, one measure is the number of times the word *hinnēh* occurs; another is the number of times a verb is followed by a noun. They then divide the book into four sections (chaps. 1–4, 5–8, 9–11, and 12–14) and evaluate each of the thirty-six criteria for each of the four sections. For each criterion, the authors statistically test whether the values of the measure for sections 3 and 4 differ significantly from the values for sections 1 and 2. Next they try to ascertain if the thirty-six measures are statistically independent, and, failing to reject independence, they use a standard statistical procedure to combine the thirty-six individual tests. The authors also carry out a similar test to determine whether section 4 differs from sections 1, 2, and 3 combined.

The most serious statistical errors occur in testing the independence of the thirty-six measurements. The first error concerns the use of the correlation coefficient (computed between each pair of criteria). Correlation coefficients are properly introduced only when there are two series of measurements and when each series has a *common* mean value. In the present case, the two series are the four values for one criterion and the four values for another. The main statistical question concerns the equality of the means of the measures for different sections, and this equality is eventually rejected. Thus, the assumption of equality of means is completely inappropriate in this context. If the means are not equal, the distribution of r_{ij} (the correlation between *ith* and *jth* criterion) will not be uniform (as claimed on p. 46) but will tend to be more concentrated near zero. The actual distribution would be quite complicated and would depend strongly on the actual values of the four different means.

A more fundamental error is the use of the chi-square test.[4] The chi-square test assumes that the observations being counted in each category are statistically independent. Clearly the 630 r_{ij} values are *not* independent: for example if r_{12} and r_{13} are large, r_{23} must also be large. Thus, the use of the chi-square test here is simply incorrect.

Perhaps the most serious error concerning the test for independence involves general statistical methodology. Although the assumption of statistical independence of measurements is commonly made, statisticians make this assumption only when they are reasonably certain that the measurements were indeed taken independently. If the statistical independence of the measurements is not clear a priori, it is incorrect to test for independence and to assume it if the test does not reject such a contention. If the measurements may be dependent, one must either use

[4] R–W, "Unity of Zechariah," 46.

an appropriate test that is not based on the assumption of independence (see section III), or one must carefully assess the effects of dependence and properly adjust for them.

In the present case it is clear that the measurements are in fact dependent. The proportions of words with k phonemes must sum (over k) to unity, and this implies that these measurements are dependent (i.e., for example, in table 15, measurement 34 is determined exactly from measurements 24–33). Even more dependence appears in the word transition frequencies (measurements 10–23). For example, the number of transitions from "verb" to some other category must equal the number of transitions to "verb" from some other category (unless the section begins with a verb, in which case there will be one additional transition from verb to something else). Thus, the sum of measurements 14, 16, and 17 must equal that of 11, 19, and 22; and similarly for the other word categories. Since the authors have deleted the EN transitions from table 15, there are actually only three exact linear restrictions and one inequality imposed in this way. In addition, the sum of all transitions must be exactly one less than the number of words (plus stops); this imposes one more linear restriction. These linear restrictions imply that the measurements are definitely correlated and thus require the use of multivariate techniques that incorporate this correlation (see section III for a discussion of one appropriate methodology).

There is also a statistical error made in the construction of the t-test.[5] First, there is what appears to be a typographical error: the plus signs in the denominator of t_k' should be minus signs. More important, the t_k' statistic as indicated is appropriate *only* under a statistical model that assumes that the means of the third and fourth sections are the same (i.e., the means for chaps. 9–11 and 12–14 are the same). Since the following test on p. 49 indicates that chaps. 9–11 and 12–14 are different, this assumption should not be made. If a t-test were to be used, it should simply test chaps. 1–8 (the first 2 sections) against 9–11 (or against 12–14 separately). Thus, (using the notation on p. 48) a statistic of the form

$$t_k^* = (.5(x_{k11} + x_{k12}) - x_{k21}) \,/\, \sqrt{.75(x_{k11} - x_{k12})^2}$$

should be used. However, since the measurements are not independent (as indicated above), the different t_k^* tests should not be combined anyway.

Finally, there is a minor problem with the numerical values listed for the word transitions (measurements 10–23): the EN transitions were deleted (these values could be determined from the other values using the linear equation discussed above), and the counts were normalized by dividing by the number of words—and not (as is more appropriate) by the number of words plus stops. However, neither of these is actually an error.

[5] R-W, "Unity of Zechariah," 48.

III. Statistical Methodology for Authorship Studies

The basic methodology for the statistical assessment of authorship involves three steps: (1) Numerical measures that are designed to reflect the unique style of a specific author are defined. (2) A statistical model for the distribution of these measures over the collection of all works by a specific author is introduced. Some scholars use the Central Limit Theorem to justify the Normal distribution as a reasonably approximate model (so R–W), but other models may be appropriate.[6] If the body of literature by a specific author is large enough, it may be possible to use "non-parametric" models which make minimal assumptions. (3) The observed distribution of each measure over the known body of works attributed to a specific author is statistically compared (in terms of the assumed model) with the observed distribution over the body of disputed works. That is, if the values of a measure over a number of disputed works show a distribution (significantly) different from that of the values over a number of known works, then one has statistical evidence of different authorship.

Step 3 emphasizes the necessity of having a reasonably large body of works known to be written by a specific author. Even if the normal model is assumed, a sufficiently large number of observations of each measure is needed to estimate adequately both the mean and variance. If the measurements are not independent, further observations are needed to estimate correlations. In the present case, there are only about 1,750 words (chaps. 1–8) attributed unequivocally to "Zechariah." R–W break this material into two sections, thus providing the minimal number of observations needed to estimate the mean and variance. Additional observations are necessary to estimate correlations between the measures. These could be obtained by subdividing the material into smaller units (e.g., into individual chapters or into groups of 250 words), but it is not clear that the range of any measure over so small a body of writing is actually indicative of the total variation of which the author is capable. This problem is exacerbated by the fact that the first eight chapters are prose while the last six are almost entirely poetry; and, hence, statistical measures may be quite differently distributed even for the same author.

The basic point is this: even a highly statistically significant difference in the distribution of measures between the first eight chapters and the last six chapters may not indicate different authorship. The first eight chapters may simply be too restricted a sample to indicate the true range of variation of one author. Conversely, failure to detect significant differences does not indicate a single author: there could certainly be two (or more) authors

[6] F. Mosteller and D. L. Wallace (*Inference and Disputed Authorship: The Federalist Papers* [Reading, MA: Addison-Wesley, 1964]) use a negative binomial model for specific word frequencies.

whose styles are indistinguishable in terms of the particular measures used.

If one is willing to subdivide the book into a larger number of smaller pieces (as indicated above), there are two reasonable methodologies that could be used. The first assumes that the measurements are normally distributed but not independent. In this case, multivariate techniques can be applied: in particular a "multivariate analysis of variance" could be carried out to test for differences among chaps. 1–8, 9–11, and 12–14.[7] This methodology would avoid the unwarranted assumption of independence among the measurements. Related multivariate analyses are also possible. A second methodology would be to apply the techniques discussed by Mosteller and Wallace (see n. 6) to the specific word counts, perhaps introducing additional words.

If, however, one is willing to make a somewhat rash statistical assumption, it is possible to apply the powerful method of chi-square tests to much of the data listed by R–W. In particular, one may consider the following three sets of frequency counts for each of the four sections of the book (chaps. 1–4, 5–8, 9–11, and 12–14):[8] (1) specific word frequencies (measurements 1–5); (2) word type transitions (measurements 10–23, multiplied by the number of words, and the number of EN transitions); (3) frequencies of "phonemes per word" (measurements 24–34, multiplied by the number of words).

One statistical model would treat each of these three data sets as a contingency table (with special structure imposed on the second one). To be specific, for each section of the book and each data set, one would consider each measurement as a frequency of observations in a category. Thus, there are five categories in the first data set. Then one could assume that each section of the book is characterized by a fixed vector of proportions for the categories defined by the measurements in each data set. This model assumes that the proportions do not vary over the works of a single author, an assumption that can be examined by testing equality of the proportions in the first two sections. Although this process would be subject to a criticism similar to that previously made against the assumption in R–W of independence of measurements, it may still be instructive first to test equality of the first two sections and, if that test fails to reject, to make the following tests: (1) equality of all four sections (which would be indicative of only one author for the whole book); (2) equality of section 3 with the combination of sections 1 and 2; (3) equality of sections 1 and 2 and 3; (4) equality of section 4 with the combination of sections 1 and 2; (5) equality of section 4 with sections 1, 2, and 3 combined (if test 2 fails to

[7] See, for example, T. Anderson, *Introduction to Multivariate Statistical Inference* (New York: John Wiley & Sons, 1958) chap. 8.

[8] Values taken from R–W ("Unity of Zechariah," 44–45) reexpressed as counts instead of proportions.

reject); (6) equality of sections 3 and 4.

Results of the standard chi-square computations for contingency tables are listed in table 1, with values significant at level .01 starred.[9] The second data set (transition counts) shows inequality for sections 1 and 2; and, hence, this data set should be ignored. The first data set (specific word frequencies) shows clearly that sections 3 (chaps. 9–11) and 4 (chaps. 12–14) are different from sections 1 and 2 (chaps. 1–8) and from each other. The last data set (phoneme frequency distribution) fails to distinguish section 3 from 1 and 2.

TABLE 1

Chi-square values (degrees of freedom) for tests of hypotheses for three data sets

	Data Set (Measurement Numbers)		
	Specific Words	Transition Frequencies	Phoneme Frequency Distribution
Test	(1–5)	(10–23)(with EN)	(24–34)
1 = 2	5.13 (4)	24.89* (10)	8.45 (10)
1 = 2 = 3 = 4	64.89* (12)	–	67.80* (30)
1 + 2 = 3	16.91* (4)	–	14.00 (10)
1 = 2 = 3	22.31* (8)	–	23.16 (20)
1 + 2 = 4	26.51* (4)	–	35.54* (10)
1 + 2 + 3 = 4	40.15* (4)	–	46.75* (10)
3 = 4	39.87* (4)	–	46.25* (10)

*Significant at level .01

The putative conclusion would be that there are three different authors, but the facts that the statistical model does not permit variation over the works of a single author and that there is a paucity of material from any one author require a far more cautious conclusion: the third and fourth sections show significant differences in the frequency of use of certain words, and these differences are not exhibited between sections 1 and 2. Also the "phoneme" distribution in the fourth section differs from that in the first three. Statistical procedures based on the previously described more general models may provide stronger evidence of multiple authorship than the above analysis, but it would be highly unreasonable to expect conclusive statistical evidence in this particular case.

[9] For a description of this methodology, see Y. Bishop, S. Fienberg, and P. Holland, *Discrete Multivariate Analysis* (Cambridge, MA: MIT Press, 1975).

IV. Methodological Integration

As has just been suggested, a statistical analysis of the unity of Zechariah indicates the presence of three distinct units: Zechariah 1–8; 9–11; 12–14. At this point one could end the discussion with guarded statements about the difficulty of achieving firm conclusions on the basis of so limited a body of textual material. It is, however, possible to continue the analysis of the book's unity by comparing the statistical results with the results from investigation by other methodological perspectives. If the results from other methods corroborate the findings achieved through statistical analysis, then what was tentatively affirmed on purely statistical grounds may be construed as a much more probable conclusion on the basis of an integrated methodological perspective.

Although a definitive study of the unity of the book of Zechariah is obviously impossible in an article, it is possible to compare succinctly the results of other methods with those gained from statistical analysis. And, in fact, such comparison confirms the statistical conclusions we have just presented. To demonstrate this point, we call attention to the following areas of analysis: content, form criticism, and tradition history.[10]

First, simply in terms of content there are obvious differences between 1–8 and 9–14.[11] In Zechariah 1–8, specific personages—especially Zechariah and Joshua—play prominent roles. No such names occur in 9–11 or 12–14. So too, specific dates are important in 1–8, and such chronological material is wanting in the final chapters. And on more general grounds, the anointed royal and the high priestly figures function importantly in Zechariah's restoration ideology as that is presented in Zechariah 1–8. In 9–11, however, we hear only of shepherds, a metaphoric reference to the community's leaders; and in 12–14, there is only vague reference to such community functionaries. Similarly, the temple is obviously a matter of great concern in Zechariah 1–8, especially in the so-called oracular material. In 9–14, however, the temple, though it is mentioned (Zech

[10] It is, of course, the case that scholars have used these and other methods to argue that Zechariah 9–11 and 12–14 are not homogeneous blocks of material. M. Saebo maintains that Zechariah 9–14 comprise four basic *Überlieferungseinheiten*: 9–10; 11; 12–13; 14 (*Sacharja 9–14, Untersuchungen von Text und Form* [WMANT 34; Neukirchen-Vluyn: Neukirchener Verlag, 1969] 313); W. Rudolph argues for three distinct blocks: Zechariah 9:1–11:3; 11:4–13:9; 14 (*Haggai—Maleachi* [KAT 13/4; Gütersloh: Gerd Mohn, 1976] 241–43); and P. Hanson identifies six basic units: Zech 9:1–17; 10:1–12; 11:1–3; 11:4–17; 13:7–9; 12:1–13:6; 14:1–21 (*The Dawn of Apocalyptic. The Historical and Sociological Roots of Jewish Apocalyptic Eschatology* [Philadelphia: Fortress, 1975] 280–401).

[11] We omit here explicit consideration of vocabulary, since that data is to a certain extent included in the statistical analysis. See, however, the classic summary of H. Mitchell (*Haggai, Zechariah, Malachi and Jonah* [Edinburgh: T. & T. Clark, 1912] 236) and the equally significant earlier work of R. Eckardt ("Der Sprachgebrauch von Zach. 9–14," *ZAW* 13 [1893] 76–109).

11:13 and 14:20–21), is more an economic institution than a locus of holiness for the restored community. From Zech 14:20–21, one senses that even though the temple exists its proper functioning remains hoped for—it is not a present reality for the writer. On grounds of content, then, there is clear warrant for distinguishing Zechariah 1–8 from 9–14 and, to a lesser extent, for distinguishing 9–11 from 12–14.

Second, form-critical considerations further buttress the conclusion that Zechariah 1–8, 9–11, and 12–14 comprise separate entities. Zechariah 1–8 consists of an elaborate interweaving of oracular and visionary material.[12] Both visions and oracles are typical of earlier prophetic discourse, though of course the particular form of the visions represents something of a middle point between the earlier prophetic visions found in Amos and the later visions present in the book of Daniel.[13] As one would expect, formulas typical for visionary and oracular discourse are prominent in Zechariah 1–8; so "thus says Yahweh of Hosts" (Zech 1:3, 14, 17; 2:8; 3:7; 8:9, 14); "I lifted up my eyes and saw" (Zech 2:1, 5; 5:1; 6:1).

These formulas are, however, infrequent in the concluding six chapters, and this because forms of discourse typical in Zechariah 1–8, oracles and visions, are not present in their classical form in this latter material. Rather than "thus says Yahweh of Hosts" one hears instead the standard deutero-prophetic connector "on that day," especially in the final block, Zechariah 12–14. Instead of the standard prophetic speech forms, the rhetorical forms used in Zechariah 9–11 are divine warrior hymns (9:1–17; 10:1–12), taunt song (11:1–3), and commissioning narrative (11:4–17).[14] In this block there are speech forms drawn from Israel's cult as well as from the standard repertoire of the prophets. But these speech forms, especially those used by earlier Israelite prophets, have been radically reformulated. The commissioning narrative has become a prophecy of doom and not simply a warrant to perform as prophet. Finally, the literature in 9–11 is poetic in form, not prose as was the case with much of Zechariah 1–8.[15]

The situation is again different with Zechariah 12–14. There are two

[12] On this issue, see the standard commentaries and W. Beuken, *Haggai-Sacharja 1–8. Studien zur Überlieferungsgeschichte der frühnachexilischen Prophetie* (Studia Semitica Neerlandica, 10; Assen: Van Gorcum, 1967); A. Petitjean, *Les oracles du proto-Zacharie. Un programme de restauration pour la communauté juive après l'exile* (Paris: Gabalda, 1969).

[13] For a recent study of the visions, see C. Jeremias, *Die Nachtgesichte des Sacharja. Untersuchungen zu ihrer Stellung im Zusammenhang der Visionsberichte im Alten Testament und zu ihrem Bildmaterial* (FRLANT 117; Göttingen: Vandenhoeck & Ruprecht, 1977).

[14] For the sake of convenience, we refer to the recent analysis of this material in P. Hanson, *Dawn of Apocalyptic*.

[15] Of course, the poetry in Zechariah is unusual (see Hanson, *Dawn of Apocalyptic*, 341).

collections, Zech 12:1–13:6 and 14:1–21, both of which defy simple form-critical classification. Both units contain motifs drawn from earlier prophetic literature, creating a quilt-like composition unified by "on that day" thread. Further, this material is essentially prose, whereas Zechariah 9–11 is essentially poetry.

In sum, whether we consider the issue of poetry versus prose, the presence of formulas or the character of the genre used in the book of Zechariah, in each case there is warrant for discerning three distinct sections, the same as those delineated by statistical analysis.

Third, on grounds of tradition history one may argue for three distinct types of material. Zechariah 1–8, along with the book of Haggai, represents what may be termed the last example of classical Israelite prophecy.[16] The social location of that which is normally thought to be prophetic behavior, i.e., monarchic society, was, after Haggai and Zechariah, clearly a thing of the past. Hopes for investing Zerubbabel with royal authority along with the drive to rebuild the temple enabled Haggai and Zechariah to operate in the classic prophetic role. Within this particular setting, Zechariah seems to represent the interests of those who helped to reestablish the Judahite community, and he did this with special attention to priestly and ritual matters.

Things change markedly with Zechariah 9–11. These chapters represent what Hanson has termed early apocalyptic. The speech forms, poetic style, and ideological complex—still rather close to prophetic eschatology—depict Yahweh as active in the arena of history, though in less easily observable ways than in Zechariah 1–8. Further, in Zechariah 9–11 we sense, as we do in Trito-Isaiah, significant inner communal strife. In Zechariah 12–14, however, such intracommunal enmity escalates. In addition there is pronounced emphasis on the earth-shaking events that will accompany Yahweh's theophany and his judgment upon the evil and his blessings for the righteous. Because of the ways in which Zechariah 12–14 differs from Zechariah 9–11, it is reasonable to maintain that the latter block of material is more fully developed apocalyptic literature.

Put in the simplest possible way, one may label Zechariah 1–8 classical prophecy; 9–11 early apocalyptic; 12–14 middle and late apocalyptic. As this traditio-historical examination suggests, the delineation of three blocks of material in the book of Zechariah that is derived from statistical analysis receives striking confirmation from yet another methodological perspective.

[16] See D. Petersen, *Late Israelite Prophecy. Studies in Deutero-Prophetic Literature and in Chronicles* (SBLMS 23; Missoula, MT: Scholars Press, 1977) 1–8.

V. Summary

The results of analysis of content, form criticism, and tradition history are consistent with the results of our statistical analysis of Zechariah. One may, with good warrant, think of three distinct units in this book: 1–8, 9–11, 12–14. In this case, then, what one methodological perspective can offer only as a tentative conclusion may, when integrated with the results of inquiries according to other methods, serve as important data in constructing a comprehensive conclusion.

Our concluding recommendations are these: (1) Statistical analysis for assessing similarities and differences in literary style should adhere to the highest standards of that method. (2) Results from statistical analysis are most valuable when they are integrated into a complete picture, one created by using other colors from the methodological palette. Only with such integration can methodological particularism be avoided, and only then can the combined input from different methods provide the fullest possible explanation of the textual material under investigation.