Analyzing Patterns of Staining in Immunohistochemical Studies: Application to a Study of Prostate Cancer Recurrence

Ruth Etzioni,1 Sarah Hawley,1 Dean Billheimer,3 Lawrence D. True,2 and Beatrice Knudsen1

1Fred Hutchinson Cancer Research Center; 2Department of Pathology, University of Washington, Seattle, Washington; and 3Division of Biostatistics, Vanderbilt-Ingram Cancer Center, Nashville, Tennessee

Abstract

Background: Immunohistochemical studies use antibodies to stain tissues with the goal of quantifying protein expression. However, protein expression is often heterogeneous resulting in variable degrees and patterns of staining. This problem is particularly acute in prostate cancer, where tumors are infiltrative and heterogeneous in nature. In this article, we introduce analytic approaches that explicitly consider both the frequency and intensity of tissue staining.

Methods: Compositional data analysis is a technique used to analyze vectors of unit-sum proportions, such as those obtained from soil sample studies or species abundance surveys. We summarized specimen staining patterns by the proportion of cells staining at mild, moderate, and intense levels and used compositional data analysis to summarize and compare the resulting staining profiles.

Results: In a study of Syndecan-1 staining patterns among 44 localized prostate cancer cases with Gleason score 7, disease, compositional data analysis did not detect a statistically significant difference between the staining patterns in recurrent (n = 22) versus nonrecurrent (n = 22) patients. Results indicated only modest increases in the proportion of cells staining at a moderate intensity in the recurrent group. In contrast, an analysis that compared quantitative scores across groups indicated a (borderline) significant increase in staining in the recurrent group (P = 0.05, t test).

Conclusions: Compositional data analysis offers a novel analytic approach for immunohistochemical studies, providing greater insight into differences in staining patterns between groups, but possibly lower statistical power than existing, score-based methods. When appropriate, we recommend conducting a compositional data analysis in addition to a standard score-based analysis. (Cancer Epidemiol Biomarkers Prev 2005;14(5):1040–6)

Background

Immunohistochemical studies are designed to detect the expression of proteins at the cellular level. In such studies, an antibody is applied to a tissue specimen, and antibody binding is detected using a chromogenic substrate that results in a color change where the protein is localized. The specimen is then examined under a microscope and the extent and intensity of color staining is assessed by an observer.

Immunohistochemical studies have become an integral part of the process of biomarker development, where they are used to determine whether overexpression or underexpression of potential markers predicts disease status. For example, several novel proteins have been shown to be markedly overexpressed in prostate cancer tissue, including α-methylacyl-CoA Racemase (1), EZH2 (2), and pim-1 kinase and hepsin (3). cDNA microarray studies typically yield many candidate genes that seem to differentiate between disease groups at the RNA level. The transformation from staining profile to score is many to one. Consider, for example, the two specimen profiles in Table 1. The first profile shows increased mild to moderate staining, whereas others show very intense staining. Because of variability, like that illustrated in Fig. 1, specimens are often not coded as positive or negative for a particular antibody; rather, a summary score is calculated (1, 4). First, the percent of cells staining at a fixed number of intensity levels (e.g. none, faint, moderate, strong) is determined. We refer to this vector of percents as the staining profile associated with a specific specimen. Then the score is given as the sum over intensity levels of the percent staining multiplied by an ordinal value corresponding to the intensity level (0 = no staining, 1 = mild staining, etc.). With four intensity levels, the resulting score ranges from 0 (no staining in the entire specimen) to 300 (intense staining uniformly within the specimen). Variants of this system have been proposed; for example, instead of the percent staining at each intensity level, an integer may be recorded with values (0, 1, 2, . . . ) indicating whether the percent staining at a given level falls within a specified range (0-20%, 20-40%, 40-60%, . . . , ref. 5). Qualitative approaches, which classify specimens as positive/negative or weak-strong, have also been used (2, 3).

The transformation from staining profile to score is many to one. Consider, for example, the two specimen profiles in Table 1. The first profile shows increased mild to moderate staining relative to the second profile, which has more intense staining. However, the two sample specimens in Table 1 yield the same score (80), because the increased frequency of cells staining moderately in the first sample compensates for the small absolute increase in intense expression in the second sample. Therefore, if differences such as those represented in Table 1 are of interest, a score-based analysis may not be adequate.

This article presents an approach for directly analyzing data on the staining profiles recorded in immunohistochemical...
studies. Our goal is to provide an analytic method that takes advantage of the richness of the information in the individual profiles. We use a class of methods collectively referred to as compositional data analysis. A composition consists of a collection of probabilities which sum to 1, and which generally represent proportions of components of a mixture. Compositional data analyses are common in soil science where interest may focus on the various components (sand, silt, and clay) that make up a sample (6, 7). Compositional data methods have also been applied in biodiversity studies where the relative frequencies of different species within a specific geographic area are of interest.

Since a tissue specimen that has been stained with an antibody represents a mixture of cells with different staining intensities, compositional data methods are applicable. We show how the compositions represented by staining profiles may be graphically represented—and statistically compared—across groups sustaining different outcomes of interest. We use these methods to analyze data from an immunohistochemical study of a potential biomarker for prostate cancer recurrence, and compare results to those obtained with the score-based approach.

Materials and Methods

Compositional Data Analysis: Basic Concepts. A composition is a set of proportions, representing components of a mixture. The elements of a composition all lie between 0 and 1 and their sum is equal to 1. In immunohistochemical studies, we can think of the data from each specimen as a composition, where the different components are the discrete staining intensity levels and the data for a given specimen consist of the proportion of cells staining at each intensity level. The goal is to compare the compositions for one group of tissues (e.g., normal tissues) with the compositions for a second group (tumor tissues). In the rest of this article, we will focus on three-part compositions because we have found them to be adequate for the immunohistochemical applications that we have considered, and because they have a convenient graphical representation.

Table 1. Different staining profiles may yield similar summary scores

<table>
<thead>
<tr>
<th>Profile 1</th>
<th>Profile 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intensity</td>
<td>0 1 2 3</td>
</tr>
<tr>
<td>Percent staining</td>
<td>30 60 10 0</td>
</tr>
</tbody>
</table>

NOTE: Both profiles have a summary score of 80.

For three-part compositions, the basic graphical tool for representing data is the ternary diagram (Fig. 2A). A ternary diagram consists of an equilateral triangle, with unit altitude and vertices corresponding to the components of the composition (e.g., 0, 1, and 2 for a three-level staining profile reflecting none-mild, moderate, and intense staining). A staining profile (a, b, and c) is represented in the figure by a point P such that the perpendicular distances from P to the edges of the triangle opposite vertices 0, 1, and 2 equal a, b, and c, respectively. Thus, staining profiles that are mostly null (no stain) would be represented by points close to vertex 0; similarly, profiles that are balanced between moderate and intense staining would be on the perpendicular line from vertex 0 to the edge connecting vertices 1 and 2.

Figure 2A shows a ternary diagram for a set of three-level staining profiles from a simulated data set consisting of 25 observations in each of two groups. In the first group, specimens tend to show a mixture of light/no staining and moderate staining as evidenced by their location along the edge of the simplex joining vertices 0 and 1. In the second group, there is less moderate staining, with a corresponding increase in light/no staining together with some intense staining. The points in the second group are said to be perturbed relative to the points in the first group. A perturbation is a composition that shifts points within the ternary diagram. The perturbation operation may be thought of as the addition operation for compositions, but the identity element (analogous to zero) is the value (1/3, 1/3, 1/3). The perturbation accounting for the difference between the medians of the two samples in Fig. 2A is (0.38, 0.05, 0.57), reflecting a shift out of level 1 (moderate) to both levels 0 (none) and 2 (intense). This type of shift is particularly difficult to detect using score-based methods, because although it reflects a real biological difference, it does not lead to a marked change in the score.

To compare staining profiles across two samples, we need to determine whether the perturbation that captures any observed between-sample differences is statistically significantly different from the identity value (1/3, 1/3, 1/3). To estimate the best-fitting perturbation, we fit a regression model to a transformation of the observed staining profiles. The transformation is called the additive log-ratio (alr) transformation, and it converts any k-part composition from a vector of observations lying between 0 and 1 to a (k - 1) vector of observations lying between $-\infty$ and $+\infty$. The regression model is then applied to the transformed data. Appendix 1 details the alr transformation and its inverse (the ialr transformation), and describes the regression model. In the absence of other covariates that may affect staining patterns, the regression...
model has only one independent variable namely, an indicator of group membership. Applying the inverse of the alr transformation to the estimated coefficient of this independent variable yields the perturbation that best captures between-sample differences.

Utilization of the alr transformation requires that all elements of the observed compositions be greater than zero. In practice this is not always the case; some specimens may have no cells staining at a given intensity level. When some of the compositions have zero elements, Aitchison (6) recommends adding a small positive constant to the zero-valued cells, but we have found that adding a small amount of normally distributed random noise produces transformed data that are more likely to satisfy the assumptions required for the regression analysis (specifically, multivariate normality of the alr-transformed observations). We detail how this is done in practice when we present our analysis of the Syndecan-1 data.

The key outputs of the regression model are (a) an estimate of the median composition among observations in the baseline group, and (b) an estimate of the perturbation that best describes the differential between the medians of the baseline group and the comparison group (6, 7). As we describe in Appendix 1, methods developed by Billheimer et al. (7) can be used to construct a 95% interval around this estimated differential. The 95% interval is represented as a

Figure 2. Staining profiles from a simulated data set consisting of two groups with 25 observations per group. A. Ternary diagram. The perturbation accounting for the shift from group 1 to group 2 is (0.38, 0.05, 0.57). Data are generated according to a bivariate normal distribution with means (0.85, 1.20; group 1) and (−1.79, 0.77; group 2), and a variance-covariance matrix that has values of 5 (diagonal elements) and 2 (off-diagonal elements). The generated observations are then transformed into points in the three-dimensional simplex using the inverse of the additive log-ratio (alr) transformation. B. Ternary diagram. The median for the first group is the same as that in A. The perturbation accounting for the shift from group 1 to group 2 is (0.04, 0.38, 0.58). Data are generated as in A, but the bivariate normal mean in the second group is (0.45, −1.25).
region in the ternary diagram; if it excludes the identity point (1/3, 1/3, 1/3) we conclude that the data indicate that there is a statistically significant difference between the two groups.

To fit the regression model and construct the 95% interval, we use WinBUGS (the WinBUGS web site can be accessed at http://www.mrc-bsu.cam.ac.uk/bugs/welcome.shtml). WinBUGS is publicly available; code to fit the regression model and construct the 95% interval is available from the authors. One of the advantages of using this framework for modeling is that it allows us to extend the regression model to the setting where multiple cores are taken from the same patient, as might, for example, be the case in a paired study (both cancerous and healthy samples analyzed within each participant). Appendix 1 describes how this may be done using WinBUGS.

Syndecan-1 and Prostate Cancer Recurrence. We previously described (4) the results of a study to investigate the association between prostate cancer progression and expression of Syndecan-1, a cell surface proteoglycan implicated in sequestering growth factors and enhancing their activity (8). One section of prostate cancer tissue from each of 76 patients diagnosed with clinically localized disease was analyzed for Syndecan-1. All tumors were Gleason score 6 or 7 and were treated with radical prostatectomy. All subjects were followed for at least 5 years. The study end point (time of recurrence) was defined as two consecutive rising serum prostate-specific antigens greater than 0.2 ng/mL. In the group with Gleason score 7 disease, 22 experienced prostate-specific antigen recurrence within 5 years, and 22 survived, recurrence free, for at least 5 years.

We analyzed the immunohistochemical data for Gleason score 7 patients using both the summary score approach and compositional data analysis. Our goal was to determine whether staining profiles differed for patients who did and did not recur within 5 years. Three categories of staining intensity (1 = no/faint staining; 2 = moderate staining; 3 = strong staining) were defined. To analyze the resulting (univariate) scores, we used a t test.

For the compositional data analysis, we first added a small amount of random noise to the cells within each staining profile that had zero values, and then renormalized the resulting compositions to sum to 1. The random noise was generated as the absolute value of a Normal(0, σ²) random variable. We considered several values for σ (i.e., 0.01, 0.015, 0.02, and 0.025). The smallest value of σ for which the resulting transformed observations satisfied the regression assumptions was 0.02, and we used this value in the analysis. Our method for dealing with zero components is similar to that of Aitchison (6), reflecting the notion that zero components represent observations that fall below some minimal detectable limit. However, rather than replacing zeros by a small, fixed number, we used a random value which we found more likely to satisfy the assumptions required for the regression analysis. We assessed sensitivity to the specific random values generated by re-running the analysis several times, with different random imputations each time, and found results across the runs to be similar.

A Simulation Study. As we have already noted, in practice it is possible that a perturbation may be very different from the identity point, but with a relatively small effect on the score. Such perturbations include shifts that reflect movement out of moderate staining to both more intense and less intense staining, and vice versa. These differences may arise in practice, for example, when comparing relatively homogeneous tissues in which the majority of cells stain moderately, with tissues that consist of a mixture of disease phenotypes with a variety of staining profiles including none or mild staining and some intense staining. Because the score is linear, simultaneous increases (or decreases) in intense and mild staining tend to cancel each other out, so these perturbations are generally not reflected in the score. As an example, Fig. 2A shows data reflecting two groups of 25 compositions each, generated from two independent distributions. The median of the first distribution is (0.35, 0.50, 0.15), and that of the second distribution (0.55, 0.10, 0.35). The resulting perturbation is (0.38, 0.05, 0.57), which reflects a lesser amount of moderate staining in the second group, with a corresponding increase in the proportion of cells staining mildly and intensely. This perturbation is quite different from (1/3, 1/3, 1/3). We conducted 100 simulations from the model generating the data in Fig. 2A, and compared the number of runs for which the compositional approach detected a difference between the two groups with the number of runs for which the standard score-based approach (t test on the scores within each group) detected a difference at a significance level of 0.05. We also conducted a second set of simulations, using a model reflected by the data in Fig. 2B. As in Fig. 2A, the first group has median (0.35, 0.50, 0.15), but the median of the second group is (0.05, 0.65, 0.30), reflecting a greater frequency of cells staining at both the moderate and intense levels.

Results

Simulation Results. Over 100 simulations from the model that generated the data in Fig. 2A, compositional data analysis identified a significant difference between groups 84% of the time, whereas a t test comparing the scores within each group identified a difference only 9% of the time at the α = 0.05 level. This example confirms that certain differences between groups, such as the one represented in Fig. 2A, will tend to be missed by the score-based approach but identified by the more general compositional analysis approach. In contrast, the difference represented in Fig. 2B was detected equally well by the compositional and score-based methods, with compositional data analysis identifying a difference in 95 of 100 simulations, and a score-based t test identifying a difference in 96 of 100 simulations, at the α = 0.05 level.

Analysis of Syndecan-1 Expression and Prostate Cancer Recurrence. Figure 3A shows the staining profiles for Gleason score 7 patients by recurrence status. Figure 3B shows a box plot for the scores within each group. A t test applied to the scores yielded a P value of 0.05 (recurrent versus non-recurrent cases).

Figure 4A shows the 95% region for the estimate of the perturbation that best describes the difference between the median compositions within each group. This region includes the identity point, suggesting lack of evidence of a significant difference between the staining profiles in the two groups. The best-fitting perturbation is (0.28, 0.41, 0.31), reflecting a modest increase in the frequency of moderate staining, but little change in the frequency of intense staining among recurrent relative to non-recurrent cancer cases. For a run of 10,000 iterations, the run time was less than 30 seconds to fit this model.

Figure 4B shows the estimated median staining profiles within each group (recurrent versus non-recurrent). The plot for the non-recurrent group shows that there is very little staining in general in this group. The figure also provides a quantitative representation of the information lost when conducting a score-based analysis, assuming scores of 1, 2, and 3 corresponding to the three levels of staining intensity represented in the figure (4). These numerical values for each staining level imply a score range with a lower limit of 100 and an upper limit of 300. The parallel lines on the figure represent points with equal score differences relative to the median.
staining profile for the non-recurrent group. Thus, for example, the points on the topmost line, which includes the observed median profile for the recurrent group, all have a score difference of 13 when compared with the non-recurrent group. Similarly, the points on the lowest line all have a difference of 100 relative to the non-recurrent group. However, the staining profiles represented by different points on the same line often represent quite different phenomena. For example, a score difference of 75 could imply that samples from recurrent patients tend to show mostly moderate staining, or a balance between negative and intense staining and very little moderate staining. As shown by our first set of simulations, information to distinguish between these possibilities would be provided by a compositional analysis.

**Discussion**

This study presents a novel application of compositional data methods to the field of immunohistochemistry. Until now, pathologists analyzing immunohistochemical data have had access to a limited set of analytic approaches. Compositional data methods extend the available analytic options, providing the ability to detect a larger range of between-specimen differences and greater insight into how staining profiles change with disease status. The ability to detect small differences between tissues as represented for example in Table 1 is particularly important, because biological behavior of a tumor may reside in a small clone of cells, which may ultimately be those that metastasize or lead to an unfavorable outcome.

Compositional data methods are designed to be applied in settings where there is a considerable amount of heterogeneity in the intensity of staining within a single specimen. In the absence of heterogeneity, the staining profiles will contain many zero components, and the assumptions of multivariate normality in the regression analysis may not be satisfied; in such settings, qualitative approaches may be preferable.

Our analyses consider the staining profile as a dependent variable and the disease status indicator as an independent variable. However, the goal of many analyses is to predict

Figure 3. Syndecan-1 staining among Gleason 7 patients by recurrence status. **A.** Ternary diagram. **B.** Box plots of the linear scores within each group.
outcome as the dependent variable, often on the basis of more than one marker. In this setting, our approach would be most useful in the exploratory phase of the analysis, to identify those markers that might be useful for predictive purposes, and to determine how staining profile differences should be quantified in the final predictor panel.

The greater flexibility provided by compositional methods may in some cases confer a cost (i.e., loss of statistical power). In the analysis of Syndecan-1 staining, we found that the compositional analysis did not identify a significant difference between the recurrent and non-recurrent groups, whereas a difference was just apparent using the more standard score-based analysis (t-test on the scores within each group; \( P = 0.05 \)). Given the limitations of the different approaches, it may be a useful practical strategy to implement both score-based and compositional analyses on any particular data set. At the very least, the compositional data approach will be a useful complement to the score analysis, allowing the investigator to confirm any results obtained and providing greater detail regarding any staining profile differences that may exist.

**Figure 4.** Results from compositional analysis of Syndecan-1 staining among Gleason 7 patients by recurrence status. **A.** Estimated perturbation describing the difference between groups together with a 95% region, which includes the identity point (1/3, 1/3, 1/3). **B.** Group medians together with lines showing points with the same linear score difference relative to non-recurrent group.
In conclusion, compositional data analysis provides a novel, flexible, and intuitive analytic approach that accounts for heterogeneity in specimen staining profiles. We anticipate that it will prove useful in many studies that aim to differentiate between patient groups on the basis of biomarker staining patterns.

Appendix 1 A statistical model for compositional data analysis

In comparing staining profiles across samples, the goal is to determine whether the perturbation that captures any observed between-sample differences is statistically significantly different from the null value.

Given an observation in the simplex, $X = (X_1, X_2, \ldots , X_K)$, and a perturbation, $a = (a_1, a_2, \ldots , a_K)$, the perturbed composition is given by $(X_{1i}a_1 / D, j = 1, \ldots , K$, where $D = \Sigma_{j=1}^{K} X_j a_j$. The null perturbation is given by $(1, 1, \ldots , 1) / K$.

To estimate the best-fitting perturbation, we first transform the compositions from the simplex to \( \mathbb{R}^{K-1} \), using the additive log-ratio (alr) transformation, which we denote by \( \phi \). Let $Y_i = \phi(X_i)$; $Y_i = (Y_{i1}, Y_{i2}, \ldots , Y_{iK-1})$, where $Y_{ij} = \log(X_{ij} / X_{iK})$. We model $Y_i$ as multivariate normal, i.e., $Y_i \approx N(\mu_i, \Sigma)$.

To model the dependence of the staining profile on disease status, where observations are available for both normal and diseased specimens, a linear model is assumed for the means of the $Y_i$'s. Thus, we have $\mu_{ij} = \beta_0 + \beta_1Z_i$, where $Z_i$ is an indicator of disease status for observation $i$. The $\beta$'s may be estimated using any package for multivariate linear regression; we have used WinBUGS, which, together with software developed by Billheimer et al. (7), conveniently allows for the computation of a 95% region around the estimated perturbation corresponding to $\beta_1$. This perturbation, which we denote $a$, is given by the inverse of the alr (ilr) transformation applied to $\beta_1$, i.e. $a_j = \exp\{\beta_{1j}\}/(1 + \Sigma_{j=1}^{K-1} \exp\{\beta_{1j}\})$ for $j=1, \ldots , K - 1$, and $a_K = 1 - \Sigma_{j=1}^{K-1} a_j$.

The WinBUGS framework requires specification of prior distributions on the vectors $\beta_0$, $\beta_1$, and $\Sigma$. In our data analysis, we use a fairly noninformative bivariate normal prior for the $\beta$'s (mean 0, variance 1, and covariance 0.5) and a non-informative inverse Wishart prior for $(\Sigma^{-1} = W(\mathbf{I}, 4))$, where $\mathbf{I}$ is the prior variance-covariance matrix specified for the $\beta$'s. An advantage of the WinBUGS framework is that it allows us to extend the model to accommodate settings in which multiple observations have been taken from the same patient. This may be particularly important in the case of prostate cancer because of within-tumor heterogeneity. To accommodate multiple observations from patient $i$, we can specify $\mu_{ij} = \beta_0 + \beta_1Z_i$, incorporating a random effect in the model so that each individual has his own median staining profile. The effect of the disease status indicator is then measured against this individual-specific baseline. To implement this extension in WinBUGS requires specification of a hyper prior on $\beta_0$, such as $\beta \approx N(\beta_0, \sigma^2)$. Then, priors for $\beta_0$ and $\beta_1$ can be specified as described above.

References

Analyzing Patterns of Staining in Immunohistochemical Studies: Application to a Study of Prostate Cancer Recurrence

Ruth Etzioni, Sarah Hawley, Dean Billheimer, et al.


Updated version

Access the most recent version of this article at:
http://cebp.aacrjournals.org/content/14/5/1040

Cited articles

This article cites 7 articles, 1 of which you can access for free at:
http://cebp.aacrjournals.org/content/14/5/1040.full#ref-list-1

E-mail alerts

Sign up to receive free email-alerts related to this article or journal.

Reprints and Subscriptions

To order reprints of this article or to subscribe to the journal, contact the AACR Publications Department at pubs@aacr.org.

Permissions

To request permission to re-use all or part of this article, contact the AACR Publications Department at permissions@aacr.org.