Journal of Modern Applied Statistical Methods

Volume 4 | Issue 2 Article 14

11-1-2005

An Estimator Of Intervention Effect On Disease Severity

David Siev
USDA Center for Veterinary Biologics, David.Siev@aphis.usda.gov

Journal of Modern Applied Statistical Methods
November, 2005, Vol. 4, No. 2, 500-508

1538 – 9472/05/$95.00

An Estimator Of Intervention Effect On Disease Severity

David Siev
USDA Center for Veterinary Biologics

When a medical intervention prevents a dichotomous outcome, the size of its effect is often estimated with the prevented fraction. Some interventions may reduce the severity of an outcome without entirely preventing it. To quantify the effect of a severity-moderating intervention, a measure termed the mitigated fraction (*MF*) is proposed. *MF* has broad applicability, because it measures the overlap of two empirical distributions based on their stochastic ordering. It is also useful in the specific context of medical interventions, because it shares certain structural and functional features with the prevented fraction. The two measures may be applied together in a single semiparametric model with components for outcome prevention and for severity conditional on the presence of the outcome.

Key words: mitigated fraction, prevented fraction, vaccine efficacy

Introduction

When a medical intervention is intended to prevent a dichotomous outcome, such as the presence or absence of disease, an estimator known as the prevented fraction (*PF*) is commonly used to measure its effect. Vaccine efficacy, for example, is often estimated using some form of prevented fraction. Some interventions are, however, intended to reduce disease severity without entirely preventing disease. It would be valuable to have an estimator that is broadly applicable for evaluating vaccine efficacy in reducing disease severity (Mehrotra, 2004). An estimator that has proved useful in animal vaccine studies is the mitigated fraction (*MF*). The mitigated fraction is a new incarnation of an old statistic with a number of salient attributes. It is both analogous in function and homologous in structure to the prevented fraction.

David Siev acknowledges helpful comments of many colleagues, particularly B. Fergen, P. Dixon, T. Katz, D. Sweeney, J. Zimmerman. Email him at David.Siev@aphis.usda.gov.

For vaccination, *PF* is the relative decrease in the probability a vaccinate will become a case, while *MF* is the relative increase in the probability that a vaccinate's disease will be less severe than a nonvaccinate's disease. This article shows its origin, describes some of its features, and illustrates how *PF* and *MF* may be components of a nested model.

Example

A swine respiratory disease vaccine study included groups of pigs treated with either vaccine or placebo. All subjects were exposed to the pathogen and subsequently sacrificed. At postmortem examination, the extent of gross lesions in the lungs of each subject was estimated by visual approximation. Two observers independently sketched on a grid the dorsal and ventral surfaces of each of the seven lung lobes. The fraction of each lobe was taken as the average of the two surfaces and two observers. The lobe fractions were weighted (by their standard relative mass) and summed to arrive at the fraction of the lungs consisting of gross lesions. They are shown in Figure 1.

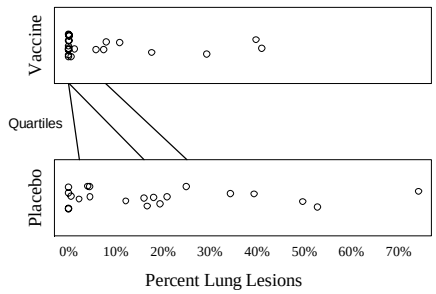


Figure 1. Fraction of lungs consisting of gross lesions. Number of subjects – placebo: 21, vaccine: 22. Points are jitter vertically to aid visualization.

How then should one analyze and summarize the findings of this study? The subjects could be divided into unaffected (0% lesions) and affected (more than 0% lesions). The prevented fraction could then be estimated, using methods for binary data. Important information is lost, however, if one only considers whether the response was present or absent and ignores its severity, particularly because most subjects were affected, and there was a wide range of response.

An approach often seen with this type of data is to calculate the average percent in each group and compare the group averages by their difference or relative difference. Taking averages is not the soundest way to summarize data that are highly skewed and border a boundary of the parameter space. The resulting summary measure also does not illuminate the vaccine's impact on individual subjects, as does *PF*, which is the relative decrease in the probability a vaccinate will become a case. A measure analogous to *PF* is *MF*, the relative increase in the probability that a vaccinate's disease will be less severe than a nonvaccinate's disease. An interesting question is whether to estimate *MF* for the entire set of data, or only for those affected by challenge. That point will be considered further when the example is revisited.

Mitigated Fraction

Prevented fraction has the general form $PF = 1 - p_2/p_1$, where, say, p_1 is the expected fraction of nonvaccinates affected by disease, and p_2 is the corresponding expectation among vaccinates. As the usual estimator of vaccine effect, *PF* is often simply termed vaccine efficacy (*VE*) in vaccine studies. Besides binomial expectations, *VE* may be constructed from other parameters that are related in some way to the probability of disease transmission (see Table 1 of Halloran et al., 1997, for an overview).

Suppose that all subjects in a vaccine trial become sick, whether vaccinated or not. Rather than looking at the effect of vaccination on the relative probability of contracting the disease, one might now wish to consider the effect of vaccination on the relative probability that the disease is milder. An estimator may be constructed that is both analogous to *PF* in function (summarizing subject probabilities) and homologous to *PF* in structure (difference relative to nonintervention).

To highlight these features, it is called the mitigated fraction (*MF*). That is, $MF = 1 - t_2/t_0$ where t_2 is the estimated probability that a vaccinate's disease is more severe than that of a nonvaccinate, and t_0 is the probability of greater severity in the absence of vaccination. *MF* may range from -1 to 1, unlike *PF*, which can take any real value no greater than 1. The difference in their ranges is related to the fact that the constituent probabilities in *MF* are relative (more or less severe than the other treatment group), while those in *PF* are not (presence or absence of disease). In practice, if a vaccine does not actually cause disease, both *MF* and *PF* will take values from 0 to 1.

If disease severity can be graded by some continuous measure or discrete assessment in a way that results in unambiguous ranks, the mitigated fraction is estimated by

$$MF = \{2W_1 - n_1(1 + n_1 + n_2)\}/n_1 n_2$$

where *W* is the familiar Wilcoxon rank sum statistic, *n* is the number of subjects in a group, and the subscripts are 1 for nonvaccinates and 2 for vaccinates.

Background

A general problem is how to distinguish between samples of two populations in some quantifiable way that avoids all parametric assumptions. A useful approach is to consider the stochastic ordering of the two empirical distributions. Figure 2 illustrates two estimators that do so,

$$T_i = \text{Prob}(Y_i > Y_j) + \tfrac{1}{2}\text{Prob}(Y_i = Y_j) .$$

For continuous random variables Prob($Y_i = Y_j$) = 0, of course, and the second term is omitted from the figure label for simplicity, but without loss of generality. If two distributions are stochastically identical, the probability that a realization from one of them is greater or lower than a realization from the other is one half. Consequently, θ_i rescales T_i to range from –1 to 1, with 0 corresponding to the null probability, ½.

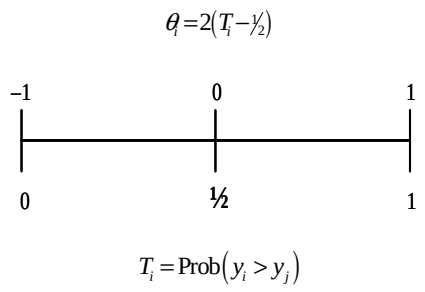


Figure 2. Because T_i and T_j are complementary probabilities, summing to one and equidistant from ½, θ_i may be reformulated as

$$\begin{aligned}\theta_i &= T_i - T_j \\ &= P(Y_j < Y_i) - P(Y_j > Y_i)\end{aligned}$$

In other words, θ_i is a measure of the overlap between the two distributions based on their stochastic ordering. A general measure of the overlap of two distributions is simply θ, the absolute value of either θ_i. $\theta = |\theta_i| = 2\left(T - \tfrac{1}{2}\right)$, where

$$T = \sup\{\text{Prob}(y_1 > y_2), \text{Prob}(y_1 < y_2)\} .$$

θ is used when comparing distributions that have no particular relative ordering. θ_i, on the other hand, is useful when the distributions arise in a particular setting that establishes an ordered relationship. For example, population 2 may be manifesting the effect of a medical intervention that is being compared to population 1, representing placebo treatment.

These estimators are generalizations of known statistics. For example, mean ridits (Bross, 1958) are T_i, and Somers' *d* statistics (Somers, 1962) are θ_i. (Vigderhous (1979) noted the connection between ridits and Somers' *d*). Somers' *d* was conceived as a measure of association between two ordinal variables, in contrast to ridit analysis, which was designed to compare the distributions of an ordinal variable in each of two distinct populations. Here, they are generalized to encompass data of all types that are not necessarily categorical and may arise from independent or correlated distributions. This general approach has been advocated by other authors (Wolf & Hogg, 1971).

It is well known that an estimate of *T* may be recovered from the Wilcoxon-Mann-Whitney statistic (Wolf & Hogg, 1971, equation 1). That may be done as follows.

$$T_i = \frac{U_i}{n_i n_j} = \frac{W_i - n_i(n_i+1)/2}{n_i n_j}$$

where

W_i = sum of the ranks in group *i* (the Wilcoxon rank sum statistic), and U_i = number of times a y_{jk} precedes a y_{ih} (the Mann-Whitney U statistic), i.e.,

$$U_i = \sum_{k=1}^{n_j}\sum_{h=1}^{n_i} \text{H}(y_{jk}, y_{ih}) ,$$

where

$$\text{H}(a,b) = 1 \text{ if } a<b;\ 0 \text{ if } a>b;\ \text{and } \tfrac{1}{2} \text{ if } a=b ,$$

and y_{ih} is the response of subject *h* (h = 1 ... n_i) in group *i* (i = 1, 2).

Substituting $\theta_i = 2(T_i - ½)$ gives

$$\theta_i = \left\{2W_i - n_i(1 + n_i + n_j)\right\} \Big/ n_i n_j$$

Stratified Design

To estimate θ from stratified data use $T_i = \sum_r U_{ir} \Big/ \sum_r n_{ir} n_{jr}$, where *r* indexes the strata. For matched pairs, this reduces to a simple binomial fraction $T_i = \sum_r \mathrm{I}(y_{jr} < y_{ir}) \Big/ R$, where *R* is the number of pairs and $\mathrm{I}(\bullet)$ is the indicator function. In that case, interval estimation can proceed by familiar methods for binomial fractions.

Subject Components

MF may be decomposed into the contribution of individual subjects. The component for a vaccinated subject *j* is $s_j = \frac{2}{n_1} \sum_{k=1}^{n_1} \mathrm{H}(y_{2j}, y_{1k}) - 1$, which is its contribution to $MF = \frac{1}{n_2} \sum_{j=1}^{n_2} s_j$. *MF* is thus the mean of the individual subject components.

Confidence Intervals

Confidence intervals using normal approximations can be derived from the asymptotic variance for *W* or the asymptotic variance for Somers' *d* provided by popular software packages. Such intervals depend on assumptions are preferably avoided and may even contain inadmissable values. An alternative is to calculate confidence intervals for *MF* by one of the bootstrap methods (Efron & Tibshirani, 1993); this is an area of ongoing investigation.

Graphical Representation (Example)

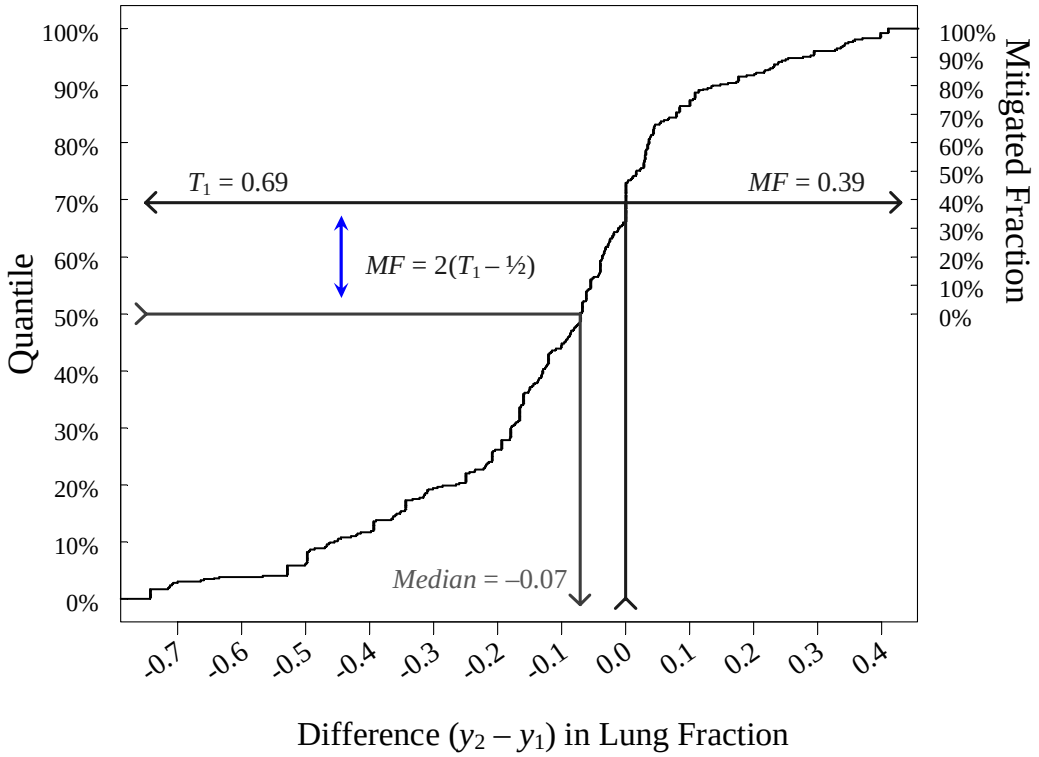
Figure 3 shows the empirical cumulative distribution function of the difference distribution, $F(Y_2 - Y_1)$, obtained from taking all pairwise differences between the groups in our example: $d_{ij} = y_{2i} - y_{1j}$, where $i = 1,...,n_2$ and $j = 1,...,n_1$. The arrow leading from the 50% quantile indicates the median difference (the Hodges-Lehmann estimator), which gives some idea of the amount of shift between the two distributions. The quantile corresponding to a difference of zero is the probability that a vaccinate's disease is less severe than that of a nonvaccinate (T_1). Rescaling the difference between T_1 and the median gives *MF*, shown in the right hand *y* axis. *MF* is thus a rescaled quantile of the difference distribution.

In contrast to the median difference, which is in the original units of measurement on the abscissa (*x* axis), *MF* reflects probabilities on the ordinate (*y* axis). In this example, $T_1 = 0.69$ means that 69% of the nonvaccinates are expected to be more severely affected than the vaccinates, $MF = 2(T_1 - ½) = 0.39$, (*95% bootstrap CI:* 0.06 to 0.68). The vaccine benefited an estimated 39% of the 50% of vaccinates who, in the absence of vaccination, would have been more severely affected than nonvaccinates.

Interpretation and application of *MF*

MF is the increase due to vaccination of the probability that a vaccinate's disease will be less severe than a nonvaccinate's disease, relative to the probability that it would have been less severe had the individual not been vaccinated. It is important to avoid direct comparison between *PF* and *MF*, which have somewhat different implications. Many of the usual estimators of vaccine efficacy are concerned with the prevention of outcomes that

Figure 3. Empirical difference distribution showing *MF* as a rescaled quantile.



are links in the chain of disease transmission, such as infection or infectivity, and in this respect *MF* is not like them. *PF* also relies on explicit case definitions, while *MF* is intended for situations where disease severity need only be clearly graded.

MF is analogous to *PF* in that it is based on estimated subject probabilities. Some relative difference measures that attempt to mimic *PF* in formulation may not necessarily have an analogous implication and should be interpreted cautiously. For example, a formulation that is often used to emulate *PF* is the relative difference of means $((\bar{y}_1 - \bar{y}_2)/\bar{y}_1)$. This is, at best, a comparison of population averages rather than subject distribution. It is rarely appropriate as the sole assessment of vaccine efficacy when the outcome is continuous rather than dichotomous (and it is particularly misleading when the data may not have arisen from a location-scale distribution). Although such estimators may be devised to emulate the configuration of *PF*, they fail to capture a similar meaning, since what is important about the constituent parameters in *PF* is not that they are means but that they are category probabilities. In this respect, *MF* is an estimator that is analogous to *PF*.

The use of mean based estimators may also arise from an understandable desire to quantify the amount of severity reduction. Unfortunately, such estimators are sensitive to the form and scale of the response measurement, which may vary substantially between similar studies. *MF*, on the other hand, is invariant to order-preserving transformations of the data. The price for such invariance is that *MF* gives no information about the magnitude of disease severity reduction, and a large value of *MF* may result from a small but highly probable reduction

in severity. That is why it is a good idea to accompany *MF* with an estimator in the original units of measurement, such as the empirical quartiles illustrated in Figure 1.

MF may also be estimated under a range of parametric assumptions, thereby offering a common approach to studies of various types. The example illustrates its most general application, where there are no assumptions other than that the data are legitimately ranked. *MF* could just as readily be estimated from ordinal categories or continuous data. With categorical data, the estimator based on *W* corresponds to the ridit estimator. In parametric analyses, the probabilities are obtained from the estimated cumulative distribution functions. For example, the frequency table shows the number of subjects of a drug trial in categories of increasing disease severity. (The data are a subset of those analyzed by Poon (2004).) By the formula, estimated *MF* = 0.08 (*95% bootstrap CI:* -0.07, 0.23). By Poon's latent normal model, estimated *MF* = 0.10 (*95% profile likelihood CI:* -0.11, 0.30). Regardless how the probabilities are estimated, the meaning of *MF* remains the same.

	increasing disease severity →				
placebo	2	22	54	29	3
drug	4	23	45	22	2

Conditional *MF* in Nested Models

Nested Model 1

Consider a model with a component for the presence or absence of disease and a component for disease severity among only those who become sick. Suppose resistance to the pathogen is dichotomous, while the immune response to vaccination among those susceptible to challenge follows some discrete or continuous distribution. Such a model may be formulated

$$f(y) = \pi^{d}\left[(1-\pi)\, f(y \mid y>0)\right]^{1-d},$$

where $d = \mathrm{I}(y=0)$ (i.e. d is an indicator taking the value 1 if y=0 and 0 otherwise) and $\pi = E(d)$, its expectation. The likelihood is then factored into a Bernoulli likelihood and a conditionally independent part which contributes to the total only for responders. This is a nested model with conditionally independent components. Since participation in the second part is conditional on crossing the hurdle of the first part, this type of nested model is sometimes termed a hurdle model (Mullahy, 1986).

If $f(y \mid y>0)$ were completely specified, say as a beta density, maximum likelihood estimation could be used to assess how the treatment groups differed with respect to prevention, conditional severity, or both. If complete specification is not warranted, *PF* may be estimated from the first part and MF_C, the conditional mitigated fraction among those affected, from the second part. To do so, let

$$p_i = 1-\pi_i$$

and

$$T_i^C = T_i \mid y_i>0,\, y_j>0\,.$$

Then,

$$PF = 1-p_2/p_1 \text{ and } MF_C = 2T_1^C - 1\,.$$

The conditionally independent nature of the nested components distinguishes the nested model from more complex mixture models. For example, continuous data with many zeros would, in some cases, be analyzed with a zero-inflated model. In contrast to a nested model, the nonresponse portion of a zero-inflated model describes a latent mixture of two populations, one which may be incapable of response and another capable of response but with response zero according to distribution $f_Y(y)$, leading to the formulation

$$f(y) = \left\{\lambda + (1-\lambda) f_Y(0)\right\}^{d}\left[(1-\lambda) f_Y(y \mid y>0)\right]^{1-d},$$

where λ is the population mixture parameter.

An example of a nested model for categorized data is the well-known continuation-ratio factorization of the multinomial likelihood

into conditionally independent binomial components. It may be parameterized $L(\underline{\pi}) \propto \prod_{j=1}^{J} \delta_j^{y_j}(1-\delta_j)^{n-r_j}$, where, for the *j*th of *J* categories, y_j is the category count, π_j is the category probability, $r_j = \sum_{k=1}^{j} y_k$ is the cumulative category count, and $n = \sum_{j=1}^{J} y_j$ is the total.

The continuation ratios are $\delta_j = \pi_j / \sum_{k=j}^{J} \pi_k$, the probability of being in category *j* given not in any previous category. Continuation-ratio models are useful for tabulated health events that occur in a natural sequence. For example, the impact of a pathogen on reproductive health may be seen by the presence of normal conception, gestation, parturition, and neonatal vigor, and a subject's inclusion at any stage depends on successfully passing the previous stage. Continuation-ratio models may also be applied to ordinal categories, such as disease severity, if they are similarly considered to be nested. In some situations they may offer an alternative to the more common cumulative probability models.

Suppose disease is categorized as absent, mild, moderate, and severe, and the counts for the two groups are arrayed in a 4 x 2 contingency table. *MF* could be estimated from the entire table, or separate estimates could be obtained for *PF* and MF_C. *PF* would be estimated from the 2 x 2 table collapsing over categories 2 through 4, while MF_C would be estimated from the 3 x 2 table that excludes the first category. A similar rationale could be applied to ranked data if each rank were thought to represent a discrete category.

Implications of Nested Model

What are the implications of the nested model for prevention and conditional severity? Suppose all nonvaccinates are sick while some vaccinates are unaffected ($p_1 = 1$, $p_2 < 1$), and disease severity is reduced among the vaccinates. *MF* is then a simple function of its components: $MF = 1 - (1 - MF_C)(1 - PF)$. Otherwise, in most practical situations where the vaccine both prevents disease ($PF > 0$) and reduces its severity among those affected ($MF_C > 0$), the relationship would be $MF < 1 - (1 - MF_C)(1 - PF)$. If the vaccine reduces disease severity among the affected but has no effect on disease prevention, although resistant individuals are found among both nonvaccinates and vaccinates ($p_1 = p_2 < 1$), the inequality reduces to $MF < MF_C$. In both latter situations, MF_C and *PF* provide illuminating information and may be examined separately from *MF*. On the other hand, in the unlikely but not impossible case that the vaccine were to prevent disease but increase severity among affected vaccinates ($MF_C < 0$), *MF* could be a useful summary which balances the benefit of prevention against the detriment of increased severity.

Nested Model 2

Nested models may also be constructed when the first component is at the end, rather than the beginning, of the disease process. For example, suppose participation in the evaluation of disease severity depends on whether or not a subject survives. The model would then be

$$f(y) = \left[f(y \mid x = 0)\, \pi \right]^{x} (1 - \pi)^{1-x},$$

where each observation consists of the pair $\{y, x\}$, *y* is the measurement of disease severity, and *x* takes the values 0 if the subject has died and 1 otherwise.

Implications of Nested Model 2

What are the implications of the nested model for severity given that a terminal outcome has not occurred? Suppose a subject dies. Is its prior disease severity relevant? There are several possibilities. For example, in an established clinical model where the severity of gross lesions predicts a possibly fatal disease, it may be valid to include the observations of all subjects, surviving or not, to assess disease severity. On the other hand, there may be no clear association between the observation and disease. Acute death may occur in response to pathogen challenge without any clinical signs at all. Retaining the observations of the dead

subjects when the severity measure is unrelated to a primary clinical outcome perpetuates an incoherent clinical model. In such cases, rank based methods are sometimes applied after assigning the dead subjects a common value greater than the maximum value of the surviving subjects. This approach treats death as simply the severest manifestation of disease, ignoring the qualitative difference between death and survival. A third position is that death is a critical event, but the prior disease severity of dead subjects is of no practical interest, leading us to exclude them from the evaluation of disease severity, but including all subjects when considering mortality. Since participation in disease severity evaluation is conditional on survival, a nested model may be constructed in which each observation consists of the pair $\{y, x\}$, where x indicates whether or not the subject has died, and y is the measurement of disease severity (nested model 2).

Example revisited

In the swine vaccine example, an estimate of the mitigated fraction is $MF = 0.39$ (*95% bootstrap CI:* 0.06 to 0.68). (The asymptotic approximation is 0.07, 0.71.) A number of subjects in the study did not succumb at all to pathogen challenge. Suppose resistance to the pathogen is dichotomous, while the immune response to vaccination among those susceptible to challenge follows some continuous distribution. The dichotomous response may be described by *PF*, and the continuous response by MF_C , the conditional mitigated fraction among those affected. *PF* and MF_C would be derived from the conditionally independent components of a hurdle model (nested model 1).

The value of nested models is that they allow simultaneous inference on two components that are conditionally independent. In the example, one would estimate *PF* by categorizing all observations as disease positive if the pathological lung fraction is greater than zero and disease negative otherwise. MF_C is then estimated using only the nonzero observations. Taking that approach, point and interval estimates are $PF = 0.21$ (-0.15, 0.49), and $MF_C = 0.42$ (0.01, 0.49). Apparently, the study is insufficient for conclusive inference on either one alone.

Conclusion

Although it is easily calculated from the Wilcoxon statistic, *MF* is aimed at estimation rather than hypothesis testing. Consequently, it helps focus attention on the clinical relevance of the outcome. Nonparametric tests are sometimes abused by those who seem to think that avoiding certain parametric assumptions also eliminates the need for forethought in study design. Care is particularly needed when observations are recorded in the form of derived ratings such as complex scoring schemes which, unlike simple grading scales, often do not preserve a clear correspondence of score with disease severity. Unless one is confident in the scores' validity when ranked, the methods shown here should not be used. Nonparametric analysis will not salvage a poorly designed scoring scheme.

Estimation requires an outcome that is quantitatively meaningful as well as clinically relevant. The study protocol should explicitly specify the outcome variable and describe how it will be recorded. Outcome specification should also aim to highlight the random structure of the data rather than conceal or ignore it by appeal to rank based methods.

For this reason, the use of nonparametric techniques in pivotal confirmatory studies has been discouraged (e.g. Longford and Nelder, 1999). Critics point out that reliance on nonparametric methods may simply postpone the search for a suitable scale of measurement and clarification of its stochastic nature, which are prerequisites for planning a study able to yield informative estimates of the size and uncertainty of relevant effects. Full distributional specification of a germane response variable is certainly ideal. Nevertheless, the basis of *MF* on ranks gives it the very qualities that are valuable in certain types of studies, particularly where a measure based on subject probabilities is preferable to an alternative measure formed from averages.

Because the mitigated fraction is comparable in structure and function to the prevented fraction, it is a useful method of estimating the benefit of an intervention that

reduces disease severity. Like *PF*, *MF* evaluates the intervention's effect by the probability a subject will benefit from the intervention. For this reason, MF_C and *PF* may illuminate different aspects of the same intervention when they are components of a nested model, and *MF* may be useful in comparisons between studies. For example, animal vaccine studies typically entail challenging all subjects with the virulent pathogen. The response to challenge often varies in magnitude between studies, and, when the response is an uncategorized measure of disease severity, the relative difference between mean group responses often varies, as well. While it is difficult to completely standardize the evaluation of such studies, *MF* estimates the probability of a beneficial response to vaccination, offering a way to assess the degree of vaccine effect at different times or locations.

References

Bross, I. D. J. (1958). How to use ridit analysis. *Biometrics, 14,* 18–38.

Efron, B. & Tibshirani, R. J. (1993). *An introduction to the bootstrap.* Chapman & Hall: New York.

Halloran, M. E., Struchiner, C. J., & Longini, I. M. (1997). Study designs for evaluating different efficacy and effectiveness aspects of vaccines. *American Journal of Epidemiology, 146,* 789–803.

Longford, N. T. & Nelder, J. A. (1999). Statistics versus statistical science in the regulatory process. *Statistics in Medicine, 18,* 2311-2320.

Mehrotra, D. V. (2004). Vaccine clinical trials: A statistical primer. *Biopharmaceutical Report, 12*(1), 1-7.

Mullahy, J. (1986). Specification and testing of some modified count data models. *Journal of Econometrics, 33,* 341–365.

Somers, R. H. (1962). A new asymmetric measure of association for ordinal variables. *American Sociological Review, 27,* 799–811.

Vigderhous, G. (1979). Equivalence between ordinal measures of association and tests of significant differences between samples. *Quality and Quantity, 13,* 187–201.

Wolfe, D. A. & Hogg, R. V. (1971). On constructing statistics and reporting data. *American Statistician, 25*(4), 27-30