Subject-Treatment Interaction

Gary L. Gadbury
Department of Statistics, Kansas State University, Manhattan, Kansas, U.S.A.

INTRODUCTION

That the effect of a treatment will vary among subjects is not surprising, nor is it a recent concept. Roses[1] provided an 1892 quote by Sir William Osler, "If it were not for the great variability among individuals medicine might as well be a science and not an art." Subject-treatment (S-T) interaction is, as the term implies, an interaction of specific subjects with applied treatment(s). The result of such interaction is a variability of "individual treatment effects" or "individual treatment heterogeneity" in a population of interest. Effects of treatment may be assessed as a measurement of efficacy or toxicity because variability in either may have important consequences. Although such variability has often been acknowledged as an important consideration, medicine today generally makes use of statistical information gathered about the general population (often about the "average" subject) and then applies it to the individual.[2] One exception is the interest in individual bioequivalence for drug "switchability" (cf., Ref. [3]) where effects of subject by formulation interaction are investigated, typically in crossover designs (e.g., Refs. [4–6]). S-T interaction in a crossover design is a topic considered in a later section.

When there is a high degree of S-T interaction in a population, there may be a nonnegligible proportion of the population responding differently to a treatment, and possibly in the opposite direction, from the average subject. Even within a carefully designed clinical trial, the validity of results may be compromised if there is a wide variability of individual treatment effects, prompting at least one author to suggest that inference about the mean treatment effect be supplemented with information about suspected variability of effects.[7] Variability in magnitude has been called a noncrossover interaction as long as the direction of the effect is the same across subjects, and variation in direction of individual effects has been termed a crossover interaction.[8] The latter is usually of more concern to researchers, possibly playing a role in adverse drug reactions.[9]

Advances in the field of pharmacogenomics have raised hopes that genetic information may be used to identify subjects who will "succeed" with a specific treatment.[10] Recent high-throughput technologies such as microarrays have enhanced the ability to search for such genetic contributors.[11,12] Gene-treatment interactions are only one component of S-T interaction. Other components may relate to sex, age, environment, other medications, etc.;[13,14] thus S-T interaction is an upper bound for gene-treatment interactions.[15] How much S-T interaction is explained by gene-treatment interactions is a nontrivial problem as clinical studies are often not designed to evaluate S-T interaction.[15]

Keywords: Causation; Clinical trials; Individual effects; Potential outcomes; Treatment heterogeneity.

The degree of S-T interaction in a population can be quantified by a parameter that is related to the variance of individual treatment effects, although this parameter cannot be directly estimated using observable data.[16] Sometimes a comparison between a proportion of subjects benefiting on a new treatment and the proportion benefiting on a standard treatment is interpreted as being related to the probability that an individual subject will benefit on the new treatment. This is not quite true. Even with crossover trials, identification of S-T interaction depends on how one defines a "true individual treatment effect." However, S-T interaction can produce effects that are testable using such data. Methods to detect these effects have focused on identifying subsets or covariates to explain perceived treatment heterogeneity.[8,17–20]

Kent and Hayward[21] comment that "baseline risk" across subjects in a study may be highly skewed, thus nearly ensuring that the summary results (e.g., estimates of average effects) will be different from the typical patient. They, as well as Kent et al.,[22] suggest that baseline risk and competing risk heterogeneity may explain treatment effect heterogeneity, and stratifying on metrics derived from these risks can enhance detection of subgroups that respond differently to treatments. They propose a risk stratified analysis (versus a conventional subgroup analysis) for detecting important differences in treatment effects, and stress the importance of understanding the complex relationship among baseline risk, treatment related harm, and competing risk when making good individual-patient recommendations and decisions.

One method to detect the presence of S-T interaction is to examine the proportion of similar responses between the treatment outcome (say, X) and the control outcome (say, Y) This overlap has been called the proportion of similar responses (PSR).[23] Given two density functions $f(x)$ and $g(x)$, the PSR is defined as the area under the smaller of the two population density functions and is expressed as

$$PSR\,(f,g) = \int \min(f(x), g(x))\,dx.$$

Stine and Heyse[24] have developed a non-parametric estimate of the PSR by using kernel density estimators to estimate both $f(x)$ and $g(x)$. Note that the kernel density estimates are made on the marginal densities of the two outcome variables without consideration of the joint distribution. As such, the value of the (non-estimable) correlation parameter, ρ_{xy}, cannot be included in the calculation of the PSR. Whether the marginal distributions and, thus, the degree of overlap between them (i.e., PSR) applies to the population of interest is another matter because patients in clinical trials are often not representative of the target populations.[25] The issue of sample selection in a clinical study is not directly considered herein, nor are the various issues that arise in clinical trials such as

compliance and dropout that pose challenges for even estimating average effects (c.f., Ref. [26]).

OVERVIEW

The focus of this entry is S-T interaction as a variability in the magnitude of true individual effects of a treatment and, when such variability is large, a practical consequence may be that the effect of treatment for some individuals has important differences from the average effect. This focus is consistent with that used in earlier papers.[16,27–29] The goals of the entry are the following: Define a true individual treatment effect; quantify S-T interaction as a nonestimable population parameter; quantify the probability that an "individual" will experience an "unfavorable" effect of treatment and show that this probability is a function of S-T interaction (hence also nonestimable); show some connections between S-T interaction and causation; demonstrate what can and cannot be learned about S-T interaction using observable data; and finally, suggest design modifications or extensions that may facilitate a more precise evaluation of the magnitude of S-T interaction (herein, "nonestimable" means that observable data contain insufficient information to allow direct estimation of a parameter). Details will be provided for a two-sample completely randomized design; however, some discussion is included later for matched-pairs and crossover designs. It is hoped that the entry will at least alert the researcher to the assumptions that must be made when equating variability of "perceived" individual treatment effects with variability of true individual effects.

COMPLETELY RANDOMIZED DESIGN COMPARING TWO TREATMENTS

Individual Treatment Effects and Potential Outcomes

Suppose two treatments, T and C, are being compared in a two-sample completely randomized design. Further, suppose that at a particular point in time, the value of an outcome variable, Y, will be measured. This outcome may be a change from baseline, and it may be quantitative or dichotomous (i.e., success or failure). For a subject receiving treatment T, the outcome $Y^{(T)}$ is observable, and for a subject receiving treatment C the outcome $Y^{(C)}$ is observable. The bivariate pair $[Y^{(T)}, Y^{(C)}]$ are potential outcomes,[30,31] in that only one of the two is observable for a specific subject at a particular time.

To avoid the need for the superscripts and thus later simplifying notation, hereafter X is the outcome to treatment T and Y is the outcome to treatment C. The bivariate pair (X, Y) are potential outcomes, and the variable $D = X - Y$ defines a true individual treatment effect. Although D cannot be observed for any subject, its definition helps to conceptualize what is meant by a true individual effect and thus clarify what can and cannot be learned about these effects using observable data. The potential outcomes framework has found use in the area of statistical causality because the ultimate interest in causal inference is effects of causes (i.e., treatments) on specific subjects. (p. 947)[32] The "fundamental problem of causal inference" is that only one of the two potential outcomes, X or Y, can be observed for an individual subject.[32] One can imagine a study evaluating k treatments, and the potential outcomes would be a vector containing k outcomes (rather than two); only one of the k outcomes would be observable for a given subject at a particular time.

It is the average treatment effect, $E(D)$, that is usually of interest but it is the variance, $\mathrm{Var}(D)$, that quantifies the degree of variability of individual treatment effects and hence the magnitude of S-T interaction. S-T interaction is present when $\mathrm{Var}(D) > 0$, but this quantity cannot be directly estimated using observed data because of the fundamental problem of causal inference mentioned earlier; however, bounds for it can be derived and these can be estimated.

A Model for Continuous Outcomes and Consequences of S-T Interaction

Consider N subjects in a two-sample design and the variables X and Y represent outcomes to treatments T and C, respectively. The set of N potential outcomes has the form given below (left) that, after treatment assignment produces observed outcomes of the form shown (right), and where the "?" represents an unobservable potential outcome.

$$\begin{pmatrix} X_1 & Y_1 \\ \vdots & \vdots \\ X_N & Y_N \end{pmatrix} \underline{\text{Treatment Assignment}} \begin{pmatrix} X_1 & ? \\ ? & Y_2 \\ \vdots & \vdots \\ ? & Y_{N-1} \\ X_N & ? \end{pmatrix} \quad (1)$$

An assumption is needed that a subject's outcome will be unaffected by the treatment assignment outcome for other subjects in the study. This was termed no interference between units[33] and generalized later as a stable unit treatment value assumption (SUTVA).[34]

One could define a treatment indicator variable and develop an estimate of $E(D)$ and its standard error with respect to a finite population randomization distribution.[28,35] Here we will suppose an infinite population model, and potential outcomes (X_i, Y_i), $i = 1, \ldots, N$ are a random sample from a bivariate normal distribution with mean vector (μ_X, μ_Y), variances, ${\sigma_X}^2$ and ${\sigma_Y}^2$, and correlation ρ_{XY}. Selection bias associated with a nonprobabilistic recruitment process are beyond the scope of this entry and is discussed elsewhere.[36] After a random treatment assignment, we have the usual two-group parallel design and a mean treatment effect, $\mu_D = \mu_X - \mu_Y$, and standard error are easily estimated using observed data because they are functions of parameters in the marginal distributions of X and of Y. The quantity for S-T interaction, ${\sigma_D}^2 = \mathrm{Var}(D) = \mathrm{Var}(X - Y)$, cannot be directly estimated because $\sigma_D^2 = \sigma_X^2 + \sigma_Y^2 - 2\sigma_X\sigma_Y\rho_{XY}$ and there is no available information about ρ_{XY} in observable data; that is, it is nonestimable.

One can write ${\sigma_D}^2$ as follows:

$$\sigma_D^2 = (\sigma_X - \sigma_Y)^2 + 2\sigma_X\sigma_Y(1 - \rho_{XY}) \quad (2)$$

So there is S-T interaction present unless $\sigma_X = \sigma_Y$ and $\rho_{XY} = 1$, and the former condition can be tested using observed data but the latter cannot. Whether or not S-T interaction produces unfavorable consequences on a subset of the population depends upon how large σ_D is with respect to μ_D, a quantity that depends on ρ_{XY}. As an illustration, assume $\sigma_X = \sigma_Y = \sigma$, that there are n subjects in each of two treatment groups (so $N = 2n$), and that a positive μ_D above some threshold is a "beneficial average treatment effect" (without loss of generality, assume this threshold is zero). Letting $\tau = \mu_D/\sigma_D$, the power to detect a positive average treatment effect is $Power = 1 - T(t_{1-\alpha}, 2n - 2, \tau\sqrt{n/2})$, where $T(t_0, k, \lambda)$ is a cumulative

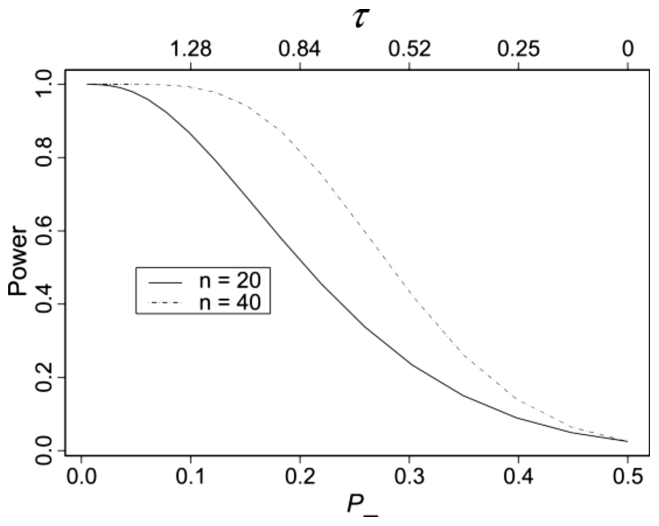

Fig. 1 Consequences of S-T interaction in a bivariate population when $\sigma_X = \sigma_Y = \sigma$, $\tau = \mu_D/\sigma$ and $\rho_{XY} = 0.7$. There are n subjects in each of the two treatment groups.

distribution function of a noncentral t random variable with k degrees of freedom and noncentrality parameter λ, evaluated at t_0. The proportion of the population experiencing a negative effect of treatment T with respect to C is $P_- = P(D < 0) = \Phi(-\tau/\sqrt{2(1-\rho_{XY})})$. Fig. 1 shows the relationship between particular ranges of τ (top horizontal axis), P_- (bottom horizontal axis), and Power (vertical axis) when the nonestimable $\rho_{XY} = 0.7$. For example, one may have high power (>80%) to detect a positive average treatment effect when more than 10% of the population experience a negative effect, unless μ_D is 1.3 times larger than σ when $n = 20$. The issue is more pronounced for larger sample sizes and for smaller values of ρ_{XY}.

In practice, one will not know the true value of ρ_{XY} and, so, will be unable to directly evaluate the size of σ_D with respect to μ_D. Little else can be done with observable data except to note that letting $\rho_{XY} = 1$ and -1 produces bounds for ${\sigma_D}^2$ that can be estimated.[16] In a data example, Gadbury, Iyer, and Allison[27] assessed the sensitivity of estimated ${\sigma_D}^2$ and its consequence, $P(D < 0)$, to a range of values for ρ_{XY} using maximum likelihood estimation. Sampling variability was assessed using large sample properties of maximum likelihood estimators (MLEs) and also using a bootstrap technique for smaller samples.

A Model for Binary Outcomes and Consequences of S-T Interaction

Gadbury, Iyer, and Albert[29] showed the following multinomial population model for binary potential outcomes:

$$\begin{array}{cccccc} (x,y) & (0,0) & (0,1) & (1,0) & (1,1) \\ P(X=x, Y=y) & \pi_1 & \pi_2 & \pi_3 & \pi_4 \end{array} \quad (3)$$

where $\Sigma_i \pi_i = 1$. Suppose that an outcome equal to 1 is a success and 0, a failure (such outcomes may also reflect a collapsing of continuous outcomes into two categories, improvement from baseline or no improvement). At a particular time, treatment T is a success with respect to C for a proportion π_3, and $D = 1$ for these subjects. However, treatment C is superior to T for a proportion π_2, and $D = -1$ for this subset. There is no difference, $D = 0$, for a proportion $\pi_1 + \pi_4$. S-T interaction is present in the population unless one of the three quantities is equal to 1: π_3, π_2, or $\pi_1 + \pi_4$. However, the individual parameters π_i, $i = 1,\ldots,4$ are nonestimable in observable data and only parameters in the marginal distributions can be estimated. The mean treatment effect is $E(D) = E(X - Y) = \pi_3 - \pi_2$ and it is this quantity that is estimated in the usual comparison of two population proportions. Suppose $E(D) = 0.4$, then it could be that $\pi_3 = 0.4$ and $\pi_2 = 0.0$, or $\pi_3 = 0.7$ and $\pi_2 = 0.3$. In the latter case, 30% of the population would be better off with treatment C although treatment T was superior on average. With no additional information, the most that can be done is to bound the individual parameters with quantities that can be estimated.[29] For example, estimable bounds for π_2 are given by the following inequalities:[29]

$$\max(0, \pi_2 - \pi_3) \le \pi_2 \le \min\{1 - (\pi_3 + \pi_4), \pi_2 + \pi_4\} \quad (4)$$

The Role of a Covariate

Suppose that a covariate, Z, is observable for all subjects in a study and that Z is unaffected by treatment. Examples are a subject's age, baseline outcome, gender, or a particular genotype. Assume that, as before, two treatments are being compared and either X or Y is observed for a subject depending on treatment assignment outcome. A well-known role for Z is to improve the efficiency in estimating a mean treatment effect. It can also be used to refine the estimable bounds for ${\sigma_D}^2$.

Suppose that (X, Y, Z) is a trivariate normal random variable and let β_{XZ} and β_{YZ} be the population regression coefficients relating Z to X and to Y, respectively. Let $\sigma_{X|Z}$ and $\sigma_{Y|Z}$ be the standard deviations of the conditional distributions of X and of Y, respectively, given Z, let $\rho_{XY|Z}$ be the partial correlation between X and Y given Z, and ${\sigma_Z}^2$ be the variance of the marginal distribution of Z. Then the following can be shown.

$$\begin{aligned} \sigma_D^2 = (\sigma_{X|Z} - \sigma_{Y|Z})^2 + (\beta_{XZ} - \beta_{YZ})^2 \sigma_Z^2 \\ + 2\sigma_{X|Z}\sigma_{Y|Z}(1 - \rho_{XY|Z}) \end{aligned} \quad (5)$$

Several points can be made regarding Eq. 5. First, all parameters in Eq. 5 can be estimated using observed data except the partial correlation. Second, if Z is a perfect linear predictor of X and of Y, then all S-T interaction is explained by a covariate-treatment interaction as $\sigma_{X|Z} = \sigma_{Y|Z} = 0$. Third, if this is not the case, then ${\sigma_D}^2 = 0$ if and only if $\sigma_{X|Z} = \sigma_{Y|Z}$, $\beta_{XZ} = \beta_{YZ}$, and $\rho_{XY|Z} = 1$, and the last equality cannot be tested using observed data. Fourth, letting $\rho_{XY|Z} = -1$ and 1 results in upper and lower bounds for ${\sigma_D}^2$ that can be estimated. Interestingly, these bounds are exactly those that result from the positive definiteness requirement of the three-dimensional correlation matrix for (X, Y, Z).[16] Gadbury and Iyer[16] developed MLEs for these bounds and derived their large sample properties. Gadbury, Iyer, and Allison[27] showed an example where data suggested a covariate by treatment interaction (i.e., $\beta_{XZ} \neq \beta_{YZ}$) and that subjects with higher baseline blood pressure appeared to have larger decreases in blood pressure after 12 weeks on a calcium supplement vs. the placebo group. They noted that although some S-T interaction is explained by $\beta_{XZ} \neq \beta_{YZ}$, there can remain individual treatment heterogeneity in subpopulations defined by values of the covariate Z. Conditioning on a value of Z, similar results for the two-sample problem (without the covariate) can now be applied to the conditional bivariate distribution of X and Y given $Z = z$, where the mean treatment effect is $\mu_{D|Z=z} = \mu_X - \mu_Y + (\beta_{XZ} - \beta_{YZ})(z - \mu_Z)$ and the conditional

variance of D is $\sigma^2_{D|Z} = \sigma^2_{X|Z} + \sigma^2_{Y|Z} - 2\sigma_{X|Z}\sigma_{Y|Z}\rho_{XY|Z}$. Gadbury, Iyer, and Allison[27] then, for varying $\rho_{XY|Z}$ between -1 and 1, assessed the sensitivity of MLEs for $\sigma_{D|Z}$ and associated MLEs for the probability of a negative treatment effect for a subpopulation of subjects defined by a value of $Z = z$, i.e., $P(D < 0 \mid Z = z)$. Large sample properties of MLEs were used to compute confidence intervals.

Working out the details of using Z as a predictor for binary outcomes and the associated refinement of bounds for π_2 is a subject for future work. Another use for Z, or any other subjective information regarding the subjects in a study, is to use this information to match subjects into pairs. Some discussion of this situation is given in the next section along with discussion of the two-period crossover design.

INDIVIDUAL TREATMENT EFFECTS IN TWO OTHER DESIGNS

Matched-pairs designs allow for observation of a treatment effect for a pair and computation of the variance of paired differences. Whether the paired differences equal true effects for individuals depends on the quality of the matching criteria. Crossover designs allow for a treatment effect to be observed for each individual because each individual receives each of the two treatments separated by a washout period. The next two subsections define individual effects and S-T interaction for these two types of designs.

Matched-Pairs Design

It is assumed that subjects are matched into pairs a priori of treatment assignment using covariate information or other subjective information thought to create homogeneity of treatment outcomes within pairs. Subjects are arbitrarily labeled within pairs as Subject 1 and Subject 2. Two treatments, T or C, are assigned within pairs via a random coin toss. There are $N = 2n$ subjects in a study resulting in n matched pairs. The potential outcomes are, again, defined using variables (X, Y) for outcomes to T and C, respectively.

Continuous Outcomes

One way to parameterize the set of $2n$ potential outcomes was shown by Gadbury.[35] This is shown below along with one pattern of observed outcomes post treatment assignment.

$$\begin{pmatrix} X_1-\varepsilon_1 & Y_1-\eta_1 \\ X_1+\varepsilon_1 & Y_1+\eta_1 \\ ---- & ---- \\ X_2-\varepsilon_2 & Y_2-\eta_2 \\ X_2+\varepsilon_2 & Y_2+\eta_2 \\ ---- & ---- \\ \vdots & \vdots \\ ---- & ---- \\ X_n-\varepsilon_n & Y_n-\eta_n \\ X_n+\varepsilon_n & Y_n-\eta_n \end{pmatrix} \underline{\quad Treatment\ Assignment \quad} \begin{pmatrix} X_1-\varepsilon_1 & ? \\ ? & Y_2+\eta_1 \\ ---- & ---- \\ ? & Y_2-\eta_2 \\ X_2+\varepsilon_2 & ? \\ ---- & ---- \\ \vdots & \vdots \\ ---- & ---- \\ ? & Y_n-\eta_n \\ X_n+\varepsilon_n & ? \end{pmatrix} \tag{6}$$

A true individual treatment effect for Subject 1 in the ith pair is $D_{i1} = X_i - Y_i - (\varepsilon_i - \eta_i)$ and for Subject 2 it is $D_{i2} = X_i - Y_i + (\varepsilon_i - \eta_i)$. The average treatment effect for the two subjects in the ith pair is $X_i - Y_i$, and the two individual effects are not equal unless $\varepsilon_i = \eta_i$. The parameters, ε_i, η_i, $i = 1,\ldots,n$, reflect the quality of the matching criteria. An observed treatment effect for the ith pair can be written as,

$$\begin{aligned} d_i = &\{X_i - Y_i - (\varepsilon_i + \eta_i)\}\delta_i \\ &+ \{X_i - Y_i + (\varepsilon_i + \eta_i)\}(1 - \delta_i) \end{aligned} \tag{7}$$

where $\delta_i = 1$ implies that Subject 1 received treatment T and Subject 2 received treatment C; if $\delta_i = 0$, the assignment is reversed for that pair and $P(\delta_i = 1) = 1/2$ for each pair. The condition, $\varepsilon_i = \eta_i = 0$, $i = 1,\ldots n$, is sufficient for equating an observed treatment effect for a pair with the true effects for the two individuals in the pair. S-T interaction can then be directly estimated using the variance of the observed paired differences.

Binary Outcomes

Gadbury, Iyer, and Albert[29] presented the issues involved in estimating π_2 in the population model shown in expression 3 using a matched-pairs design. There are now 16 outcomes in the potential outcomes framework, with four possible outcomes for each subject in a pair. The potential outcomes for Subjects 1 and 2 in a pair are represented by $\{(X_1,Y_1), (X_2,Y_2)\}$. Four (of the 16) outcomes reflect subjects who are perfectly matched on both variables. These outcomes are $\{(x_1,y_1), (x_2,y_2)\} = \{(0,0), (0,0)\}$, $\{(0,1), (0,1)\}$, $\{(1,0), (1,0)\}$, and $\{(1,1), (1,1)\}$. If the sum of the probabilities for these four outcomes is equal to 1, then only perfectly matched pairs are available in the matched population.

Suppose subjects have been randomly labeled within each of n pairs and, without loss of generality, Subject 1 in each pair receives treatment T and Subject 2 receives treatment C. So observable variables for a pair are of the form (X_1, Y_2) and resulting data are of the form for the usual matched 2×2 table for binary outcomes, or can be shown as

X_1	Y_2	Frequency
0	0	s_1
0	1	s_2
1	0	s_3
1	1	s_4

(8)

where $\Sigma_i s_i = n$. Gadbury, Iyer, and Albert 19 showed that a sufficient condition for s_i/n to be unbiased for π_i is that the population of matched subjects are matched on at least one of the two potential outcome variables; that is, the proportion that are mismatched on both X and Y equals zero. There are four potential outcomes that are a double mismatch, but exchangeability within pairs (as a result of the

random labeling) allows for these four to be expressed as two unique double mismatches: $\{(x_1,y_1), (x_2,y_2)\} = \{(0,0), (1,1)\}$ and $\{(x_1,y_1), (x_2,y_2)\} = \{(0,1), (1,0)\}$. A necessary and sufficient condition for s_i/n to be unbiased for π_i is that the probabilities of these two different occurrences of a double-mismatched pair are equal.[29]

Of course, without additional information, one will not know how well subjects are matched in the population. One can still obtain bounds for π_i similar to what was shown earlier for the two independent sample framework. The quality of the matching criteria cannot be evaluated using observed data unless the design is altered to facilitate this. For example, some pairs could be selected so that both subjects in the pair received treatment T, and other pairs where both subjects received treatment C. Gadbury, Iyer, and Albert[29] constructed a simulated example where $\pi_2 = 0.125$, and an estimated upper bound for π_2 without this design extension was 0.34. After using the extension to the design that provided information on the matching criteria, the estimated upper bound was tightened to 0.21.

An obvious extension to these results for matched pairs is to using covariates or other information on subjects to group subjects into homogeneous blocks. This design will likely be of more practical use to researchers. Evaluating S-T interaction in this type of design fora binary outcome variable was developed and reported in Albert et al.[37]

A Two-Period Crossover Design

When feasible to implement, the crossover design allows a direct observation for an individual effect as outcomes for each subject are observed for both treatments. (For more discussion on this, see Ref. [15] and for more details on crossover designs in general, see Ref. [38], or [39] for a review of papers on crossover designs.) As mentioned earlier, these designs have been used in individual bioequivalence studies. Whether or not these observed individual effects are true individual effects depends on how one defines a true individual effect. Suppose that $N = 2n$ subjects are in a trial and that during the first time period, n will be randomly selected to receive treatment T with the other n receiving C (this balance in the design is not necessary but is simply used for convenience). At some point in time (time 1), outcomes are observed. After a washout period, treatment assignments are reversed and at some second point in time (Time 2), outcomes are observed. The two sequences can be abbreviated TC (i.e., treatment T in period one and control C in period 2) and CT (control C in period 1 and treatment T in period 2). It is assumed that the washout period is sufficient so that there are no carryover effects of treatments at Time 1 to Time 2. Potential outcomes for this scenario were shown by Gadbury[35] and are repeated below.

$$\begin{array}{ccccc} \text{Subject} & \text{Time 1} & & \text{Time 2} & \\ 1 & X_1 - t_1 & Y_1 - \tau_1 & X_1 + t_1 & Y_1 + \tau_1 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 2n & X_{2n} - t_{2n} & Y_{2n} - \tau_{2n} & X_{2n} + t_{2n} & Y_{2n} + \tau_{2n} \end{array} \quad (9)$$

There is a true individual treatment effect for each subject at each time period. The parameters t, τ reflect time effects from Time 1 to Time 2. It may be reasonable to define a true individual treatment effect for the ith subject as a combination of the two effects at each time period. The average will be used here so the true treatment effect for the ith subject is $D_i = (1/2)\{(X_i - t_i) - (Y_i - \tau_i) + (X_i + t_i) - (Y_i + \tau_i)\} = X_i - Y_i$. Let $\delta_i = 1$ if the ith subject is assigned the first treatment sequence, T then C, and $\delta_i = 0$ if assigned the reverse sequence, C then T. The observed treatment effect for the ith subject is $d_i = \{(X_i - t_i) - (Y_i + \tau_i)\}\delta_i + \{(X_i + t_i) - (Y_i - \tau_i)\}(1 - \delta_i)$. A sufficient condition for this observed effect to equal the true effect is that $(t_i + \tau_i) = 0$, meaning that the sequence to which the subject was assigned had no effect on outcomes. Bounds for the variance of individual treatment effects, σ_D^2, will depend on nonestimable parameters in the distribution of time effects, particularly the variance-covariance matrix of (t,τ). If a Balaam design is used, where some subjects are allocated to receive the same treatment in both time periods (i.e., include the sequences TT and CC in the above design), then variances of the time effect parameters can be estimated but the correlation cannot.[40] Still, bounds for σ_D^2 should be estimable by letting the correlation range from -1 to 1.

CONCLUSION

This entry highlighted the role of potential outcomes for defining individual treatment effects and quantifying S-T interaction. Three designs were discussed and a few details were provided for the two-sample completely randomized design. Evaluating the degree of S-T interaction, and its consequences, is a challenging problem and often only bounds can be derived that can be directly estimated using data. The ability to estimate S-T interaction requires assumptions about the degree to which "observed individual treatment effects" are equal to true individual effects. These assumptions may be perfectly reasonable in many applications where much is known about the treatment and/or the disease that is being treated. Such clinical knowledge and past experience are often outside the realm of what data can show in a single experiment. This entry demonstrated some limitations of using data alone to evaluate individual treatment heterogeneity. It is hoped that this will help illuminate the assumptions that are made when equating observed heterogeneity with true heterogeneity and, thus, help investigators to evaluate the reasonableness of such assumptions in particular contexts. Perhaps new designs or extensions to current designs will be considered for clinical trials with an added goal in mind, as Longford remarked, "...inference about the mean treatment effect be supplemented by inference about the variation of the treatment effects". (p. 1473)[7]

ACKNOWLEDGMENTS

The author thanks Hari K. Iyer for reviewing an earlier draft of this entry and Robert Poulson and Roshan Serasinghe for assisting with updates.

REFERENCES

1. Roses, A.D. Pharmacogenetics and the practice of medicine. Science **2000**, *405*, 857–865.
2. Marshall, A. Laying the foundations for personalized medicines. Nat. Biotechnol. **1997**, *15*, 954–957.
3. Chow, S.-C.; Liu, J.-P. Preface. Stat. Med. **2000**, *19*, 2719.
4. Enrenyi, L.; Tothfalusi, L. Subject-by-formulation interaction in determinations of individual bioequivalence: bias and prevalence. Pharm. Res. **1999**, *16*, 186–190.
5. Gould, A.L. A practical approach for evaluating population and individual bioequivalence. Stat. Med. **2000**, *19*, 2721–2740.

6. Chow, S.-C.; Shao, J.; Wang, H. Individual bioequivalence testing under 2 × 3 designs. Stat. Med. **2002**, *21*, 629–648.
7. Longford, N. Selection bias and treatment heterogeneity in clinical trials. Stat. Med. **1999**, *18*, 1467–1474.
8. Gail, M.; Simon, R. Testing for qualitative interactions between treatment effects and patient subsets. Biometrics **1985**, *41*, 361–372.
9. Phillips, K.A.; Veenstra, D.L.; Oren, E.; Lee, J.K.; Sadee, W. Potential role of pharmacogenomics in reducing adverse drug reactions. JAMA **2001**, *286*, 2270–2279.
10. Evans, W.E.; Relling, M.V. Pharmacogenomics: Translating functional genomics into rational therapeutics. Science **1999**, *286*, 487–491.
11. Gunther, E.; Stone, D.J.; Gerwein, R.W.; Bento, P.; Heyes, M.P. Prediction of clinical drug efficacy by classification of drug-induced genomic expression profiles in vitro. Proc. Natl. Acad. Sci. USA **2003**, *100*, 9608–9613.
12. Allison, D. Statistical Methods for Microarray Research for Drug Target Identification. *2002 Proceedings of the American Statistical Association [CD ROM]*; American Statistical Association: Alexandria, VA, 2002.
13. Evans, W.E.; Johnson, J. Pharmacogenomics: The inherited basis for interindividual differences in drug response. Annu. Rev. Genomics Hum. Genet. **2001**, *2*, 9–39.
14. Wilson, J.F.; Weale, M.E.; Smith, A.C.; Gratrix, F.; Fletcher, B.; Thomas, M.G.; Bradman, N.; Goldstein, D.B. Population genetic structure of variable drug response. Nat. Genet. **2001**, *29*,265–269.
15. Senn, S. Individual therapy: New dawn or false dawn?. Drug Inf. J. **2001**, *35*, 1479–1494.
16. Gadbury, G.L.; Iyer, H.K. Unit-treatment interaction and its practical consequences. Biometrics **2000**, *56*, 882–885.
17. Good, P.I. Detection of a treatment effect when not all experimental subjects will respond to treatment. Biometrics **1979**, *35*, 483–489.
18. Simon, R. Patient subsets and variation in therapeutic efficacy. Br. J. Clin. Pharmacol. **1982**, *14*, 473–482.
19. Conover, W.J.; Salsburg, D.S. Locally most powerful tests for detecting treatment effects when only a subset of patients can be expected to "respond" to treatment. Biometrics **1988**, *44*,189–196.
20. Horwitz, R.I.; Singer, B.H.; Makuch, R.W.; Viscoli, C.M. Can treatment that is helpful on average be harmful to some patients? A study of the conflicting information needs of clinical inquiry and drug regulation. J. Clin. Epidemiol. **1996**, *49*, 395–400.
21. Kent D.M.; Hayward, R.A. Limitations of applying summary results of clinical trials to individual patients. JAMA **2007**, *298*, 1209–1212.
22. Kent, D.M.; Ali, A.A.; Hayward, R.A. Competing risk and heterogeneity of treatment effect in clinical trials. Trials **2008**, *9*, 30.
23. Rom, D.M.; Hwang, E. Testing for individual and population equivalence based on the proportion of similar responses. Stat. Med. **1996**, *15*, 1489–1505.
24. Stine, R.A.; Heyse, J.F. Non-parametric estimates of overlap. Stat. Med. **2001**, *20*, 215–236.
25. Senn, S. Addes values: controversies concerning randomization and additivity in clinical trials. Stat. Med. **2004**, *23*, 3729–3753.
26. McNamee, R. Intention to treat, per protocol, as treated and instrumental variable estimators given non-compliance and effect heterogeneity. Stat. Med., **2009**, *28*, 2639–2652.
27. Gadbury, G.L.; Iyer, H.K.; Allison, D. Evaluating subject-treatment interaction when comparing two treatments. J. Biopharm. Stat. **2001**, *11*, 313–333.
28. Gadbury, G.L.; Iyer, H.K. On Estimating Subject-Treatment Interaction. *Advances on Methodological and Applied Aspects of Probability and Statistics*; Taylor and Francis: New York, 2002; 349–364.
29. Gadbury, G.L.; Iyer, H.K.; Albert, J. Individual treatment effects in randomized trials with binary outcomes. J. Stat. Plan., **2004**, *121*, 163–174.
30. Neyman, J. On the application of probability theory to agricultural experiments. Essay on principles. Rocz. Nauk Rol. X (in Polish) **1923**, 1–51.—(English translation of Section 9 by D. M. T. P. Dabrowska, Speed, Stat. Sci. **1990**, *5*, 465–480).
31. Rubin, D.B. Estimating causal effects of treatments in randomized and nonrandomized studies. Br. J. Educ. Psychol. **1974**, *66*, 688–701.
32. Holland, P.W. Statistics and causal inference. J. Am. Stat. Assoc. **1986**, *81*, 945–960.
33. Cox, D.R. *The Planning of Experiments*; Wiley: New York, 1958; 19.
34. Rubin, D.B. Comment on Basu's randomization analysis of experimental data. J. Am. Stat. Assoc. **1980**, *75*, 591–593.
35. Gadbury, G. Randomization inference and bias of standard errors. Am. Stat. **2001**, *55*, 310–313.
36. Longford, N.T.; Nelder, J.A. Statistics versus statistical science in the regulatory process. Stat. Med. **1999**, *18*, 2311–2320.
37. Albert, J.M.; Gadbury, G.L.; Mascha, E.J. Assessing treatment effect heterogeneity in clinical trials with blocked binary outcomes. Biom. J. **2005**, *47*, 662–673.
38. Senn, S. *Cross-Over Trials in Clinical Research*; John Wiley and Sons Ltd.: West Sussex, 1993.
39. Senn, S. Cross-over trials in *Statistics in Medicine*: the first '25' years. Stat. Med. **2006**, *25*, 3430–3442.
40. Ndum, E. *Individual Treatment Effect Heterogeneity in Multiple Time Point Trials*, PhD Dissertation; Kansas State University: Manhattan, Kansas, 2009.