Combining Independent, Weighted *P*-Values: Achieving Computational Stability by a Systematic Expansion with Controllable Accuracy

Gelio Alves, Yi-Kuo Yu*

National Center for Biotechnology Information, National Library of Medicine, National Institutes of Health, Bethesda, Maryland, United States of America

Abstract

Given the expanding availability of scientific data and tools to analyze them, combining different assessments of the same piece of information has become increasingly important for social, biological, and even physical sciences. This task demands, to begin with, a method-independent standard, such as the *P*-value, that can be used to assess the reliability of a piece of information. Good's formula and Fisher's method combine independent *P*-values with respectively unequal and equal weights. Both approaches may be regarded as limiting instances of a general case of combining *P*-values from m groups; *P*-values within each group are weighted equally, while weight varies by group. When some of the weights become nearly degenerate, as cautioned by Good, numeric instability occurs in computation of the combined *P*-values. We deal explicitly with this difficulty by deriving a controlled expansion, in powers of differences in inverse weights, that provides both accurate statistics and stable numerics. We illustrate the utility of this systematic approach with a few examples. In addition, we also provide here an alternative derivation for the probability distribution function of the general case and show how the analytic formula obtained reduces to both Good's and Fisher's methods as special cases. A C++ program, which computes the combined *P*-values with equal numerical stability regardless of whether weights are (nearly) degenerate or not, is available for download at our group website http://www.ncbi.nlm.nih.gov/CBBresearch/Yu/downloads/CoinedPValues.html.

Introduction

Forming a single statistical significance out of multiple independent tests has been an important procedure in many scientific disciplines, including social psychology [1,2], medical research [3], genetics [4], proteomics [5], genomics [6], bioinformatics [7,8] and so on. Among the best known approaches are Fisher's method [9] and Good's formula [10]. To form a single significance assignment out of L independent tail-area probabilities, Fisher's method combines these L probabilities democratically while Good's formula weights every probability differently. Being able to weight more on better trusted *P*-values, Good's formula is versatile. Nevertheless, it suffers from numerical instabilities when weights are nearly degenerate [10]. This paper provides an analytic formula (see eq. (33)) to properly handle nearly degenerate weights. Employing complex variable theory, we have derived this controlled expansion, in powers of differences in inverse weights, that affords for the first time both accurate statistics and stable numerics.

In addition to the scenarios covered by Fisher's method and Good's formula, one may foresee the occurrence of the following *general case* (GC): independent *P*-values are categorized into groups within each of which *P*-values have the same weight, while weight varies by group. The criterion for grouping can be very general,

ranging from previously known attributes to differences in experimental protocols. As an example, one may wish to group data and their associated *P*-values by type of experimental instruments and assign each group a different weight. When there is only one instrument type, the GC reduces to Fisher's consideration. When there exist no replicates within each instrument type, the GC coincides with the consideration of Good.

In [10], Good also mentioned the possibility of obtaining an analytic expression for the GC, but did not provide it. Since Good's formula [10] contains, in the denominator, pairwise differences between weights, he cautiously remarked that his formula may become ill-conditioned when weights of similar magnitudes exist and thus calculations should be done by holding more decimal places. This statement has been paraphrased by numerous authors [11–15], and many of them have tried to seek numerically stable alternatives at the expense of using uncontrolled approximations. However, what remained elusive was a proper procedure that both provides accurate statistics and deals with nearly degenerate weights in a numerically stable manner.

The main result of this paper is an explicit formula (eq. (33)) that can properly handle nearly degenerate weights for the GC, including Good's formula of course. This derived, controlled expansion, in powers of differences in inverse weights, affords for the first time both accurate statistics and stable numerics. Employing

a complex variable integral formulation, we also provide a novel derivation of the distribution function for the GC and thus become the first, in the context of combining P-values, to make available an analytic formula for the probability distribution function for the GC.

In the statistics community, attempts to obtain an overall significance level for the results of *independent* runs of studies date back to the 1930s [9,16–18], if not earlier. Nevertheless, one should note that the mathematical underpinnings of combining P-values also appear in other areas of research. For example, the equivalent of Good's formula had emerged in 1910 in the context of sequential radioactive decay [19], while the first analytic expression for Fisher's combined P-value had emerged in 1960 as a special case of the former when all the decay constants are identical [20]. After Good's work [10], Good's formula was rederived by McGill and Gibbon [21], and later on by Likes [22]. As for the GC, Fisher's method included, the mathematical equivalents appear in different areas of studies mainly under the consideration of sum of exponential/gamma variables. The distribution functions of linear combinations of exponential/gamma variables are useful in various fields. When limited to exponential variables, it results in the Erlang distribution that is often encountered in queuing theory [23]. It is also connected to the renewal theory [24] and time series problem [25], and it can be applied to model reliability [26]. The intimate connections between these seemingly different problems are not obvious at first glance. Consequently, it is not surprising that the distribution function of the GC has been rediscovered/rederived many times and that some information about it has not been widely circulated. Our literature searches show that the first explicit result (without further derivatives involved) for the distribution function for the GC was obtained by Mathai [27]. Subsequently, motivated by different contexts, Harrison [28], Amari and Mirsa [29], and Jasiulewicz and Kordecki [26] all rederived the same distribution function.

There also exist numerical approaches for combining independent P-values. These typically involve inverting cumulative distribution functions. For example, Stouffer's z-methods [1], whether unweighted [30] or weighted [31,32], require inverting the error function. Lancaster's generalization [33,34] of Fisher's formalism also requires inverting gamma distribution function to incorporate unequal weighting for P-values combined. Since our main focus is on analytic approaches, we shall refrain from delving into any numerical method.

In the Methods section, we will first summarize Fisher's and Good's methods for combining P-values, then present the mathematical definition of the GC. In the Results section, the subsection headed by "Derivation of $\mathrm{Prob}(Q \leq \tau)$" is devoted to the derivation of the probability distribution function and cumulative probability for the GC. Since both Fisher's and Good's considerations arise as special limiting cases of the GC, we also illustrate there that our cumulative probability distribution for the GC indeed reduces to the appropriate limiting formulas upon taking appropriate parameters. In the subsection headed by "Accommodation of arbitrary weights", we delve into our main innovative part – taming the instability caused by nearly degenerate weights – and provide a formula with controllable accuracy for combining P-values. A few examples of using the main results are then provided in the Example subsection. This paper then concludes with the Discussion section. A C++ program CoinedPValues, which combines independent weighted P-values with equal numerical stability regardless of whether weights are (nearly) degenerate, is available for download at our group website: http://www.ncbi.nlm.nih.gov/CBBresearch/Yu/downloads/CoinedPValues.html.

Methods

Summary of Fisher's and Good's methods for combining P-values

Assume that a piece of information is assessed by L independent tests, each yielding a P-value. Each P-value obtained is between zero and one since, by definition, it is the probability for the experimental outcome to arise from the null model. Prior to combining these L independent P-values $(p_1, p_2, \ldots, p_L)$ to form a single significance level, we note the following. Although for any null model P-value must distribute uniformly over $(0,1]$, the L P-values obtained need not have their average close to 0.5. This is especially the case when the piece of information we are evaluating is not well described by the null model(s) considered.

For later convenience, let us define

$$\tau_F \equiv p_1 \cdot p_2 \cdots p_L , \tag{1}$$

$$\tau_G \equiv p_1^{w_1} \cdot p_2^{w_2} \cdots p_L^{w_L} , \tag{2}$$

where $w_i > 0$ is the weight associated with the ith P-value. To form a unified significance, Fisher and Good considered respectively the stochastic quantities Q_F and Q_G, defined by

$$Q_F \equiv x_1 \cdot x_2 \cdots x_L , \tag{3}$$

$$Q_G \equiv x_1^{w_1} \cdot x_2^{w_2} \cdots x_L^{w_L} , \tag{4}$$

where each x_i represents a random variable drawn from an uniform, independent distribution over $(0,1]$. The following probabilities

$$\mathrm{Prob}(Q_F \leq \tau_F) = \tau_F \sum_{l=0}^{L-1} \frac{[\ln(1/\tau_F)]^l}{l!} \tag{5}$$

$$\mathrm{Prob}(Q_G \leq \tau_G) = \sum_{l=1}^{L} \Lambda_l \, \tau_G^{1/w_l} \tag{6}$$

provide the unified statistical significances, corresponding respectively to Fisher's and Good's considerations, from combining L independent P-values. In eq. (6), the prefactor Λ_l is given by

$$\Lambda_l = \frac{w_l^{L-1}}{\Pi_{k \neq l}(w_l - w_k)} . \tag{7}$$

Apparently, Λ_l is ill-defined when the weight w_l coincides with or is numerically close to any other weights w_k. Although Fisher did not derive (5), from this point on, we shall refer to (5) as Fisher's formula and (6) as Good's formula.

General case including Fisher's and Good's formulas

Let us divide the L independent P-values into m groups with $1 \leq m \leq L$. Within each group k, we weight the n_k P-values equally; while P-values in different groups are weighted differently. Therefore, when $m = L$ and $n_k = 1$ $\forall k$, we have the Good's case; when $m = 1$ and $n_1 = L$, we reach Fisher's case. We will hence

define the following quantities of interest

$$\tau \equiv \prod_{k=1}^{m} \left[\prod_{j=1}^{n_k} p_{k;j} \right]^{w_k} , \qquad (8)$$

$$Q \equiv \prod_{k=1}^{m} \left[\prod_{j=1}^{n_k} x_{k;j} \right]^{w_k} , \qquad (9)$$

where each $x_{k;j}$ represents again a random variable drawn from an uniform, independent distribution over (0,1]. The quantity of interest $\mathrm{Prob}(Q \leq \tau)$, if obtained, should cover results of both Fisher and Good as the limiting cases. In the next section, we will start by deriving an exact expression for $\mathrm{Prob}(Q \leq \tau)$ and describing how to recover the results of Fisher and Good.

Results

Derivation of $\mathrm{Prob}(Q \leq \tau)$

Let $F(\tau) \equiv \mathrm{Prob}(Q \leq \tau)$, we may then write

$$F(\tau) = \int_0^1 \cdots \int_0^1 \theta \left(\tau - \prod_{k=1}^{m} \left[\prod_{i=1}^{n_k} x_{k;i} \right]^{w_k} \right) \prod_{k=1}^{m} \prod_{j=1}^{n_k} dx_{k;j} , \qquad (10)$$

where $\theta(x)$ is the heaviside step function, taking value 1 when $x > 0$ and value 0 when $x < 0$. Upon taking a derivative with respect to τ, we obtain

$$f(\tau) \equiv \frac{dF(\tau)}{d\tau} = \int_0^1 \cdots \int_0^1 \delta \left(\tau - \prod_{k=1}^{m} \left[\prod_{j=1}^{n_k} x_{k;j} \right]^{w_k} \right) \prod_{k=1}^{m} \prod_{j=1}^{n_k} dx_{k;j} , \qquad (11)$$

where $\delta(x)$ is Dirac's delta function that takes value 0 everywhere except at $x = 0$ and that $\forall a > 0$, $\int_{-a}^{a} \delta(x) dx = 1$.

To proceed, let us make the following change of variables

$$\tau = e^{-t}$$

$$x_{k;j} = e^{-u_{k;j}}$$

and remember that if y_0 is the only root of f $(f(y_0) = 0)$

$$\delta(f(y)) = \frac{\delta(y - y_0)}{|f'(y_0)|} ,$$

we may then rewrite (11) as

$$f(\tau) = f(e^{-t}) =$$
$$\int_0^{\infty} \cdots \int_0^{\infty} e^t e^{-\sum_{k,j} u_{k;j}} \delta \left(t - \sum_{k=1}^{m} w_k \left[\sum_{j=1}^{n_k} u_{k;j} \right] \right) \prod_{k=1}^{m} \prod_{j=1}^{n_k} du_{k;j} \qquad (12)$$

$$\equiv e^t f_e(t) . \qquad (13)$$

Note that $f_e(t)$ is exactly the probability density function of a weighted, linear sum of exponential variables.

By introducing the integral representation of the δ function

$$\delta(t - c) = \frac{1}{2\pi} \int_{-\infty}^{\infty} dq \, e^{-iq(t-c)} ,$$

we may re-express (12) as

$$f(\tau) = \int_{-\infty}^{\infty} \frac{dq}{2\pi} e^{-it(q+i)} \prod_{k=1}^{m} \left[\int_0^{\infty} e^{-u} e^{iqw_k u} du \right]^{n_k}$$

$$= \int_{-\infty}^{\infty} \frac{dq}{2\pi} e^{-it(q+i)} \prod_{k=1}^{m} \left[\frac{1}{1 - iqw_k} \right]^{n_k}$$

$$= \int_{-\infty}^{\infty} \frac{dq}{2\pi} e^{-it(q+i)} \prod_{l=1}^{m} \left(\frac{i}{w_l} \right)^{n_l} \prod_{k=1}^{m} \left[\frac{1}{q + ir_k} \right]^{n_k}$$

$$= \left[\prod_{l=1}^{m} r_l^{n_l} \right] (i)^{\sum_{k=1}^{m} n_k} \int_{-\infty}^{\infty} \frac{dq}{2\pi} e^{-it(q+i)} \prod_{k=1}^{m} \left[\frac{1}{q + ir_k} \right]^{n_k} \qquad (14)$$

$$\equiv \left[\prod_{l=1}^{m} r_l^{n_l} \right] \tilde{f}(\tau; n_1, n_2, \ldots, n_m) , \qquad (15)$$

where $r_k \equiv 1/w_k$ is introduced for the ease of analytic manipulation and $\tilde{f}$ is introduced for later convenience. Since all $w_k > 0$, implying that all $r_k > 0$, the poles of the integrand in (14) lie completely at the lower half of the q-plane. Consequently, the integral of q may be extended to enclose the lower half q-plane to result in

$$f(\tau) = e^t \left[\prod_{l=1}^{m} (ir_l)^{n_l} \right] \left(\frac{-2\pi i}{2\pi} \right)$$

$$\sum_{k=1}^{m} \frac{(\partial/\partial q)^{n_k - 1}}{(n_k - 1)!} \left[e^{-itq} \prod_{j=1, j \neq k}^{m} \left(\frac{1}{q + ir_j} \right)^{n_j} \right]_{q = -ir_k}$$

$$= e^t \left[\prod_{l=1}^{m} (ir_l)^{n_l} \right] \sum_{k=1}^{m} \left\{ (-i) \sum_{\substack{g_1, g_2, \ldots, g_m = 0 \\ \sum g_i = n_k - 1}} \frac{(-1)^{n_k - 1} (it)^{g_k}}{g_k! \, e^{r_k t}} \times \right.$$

$$\prod_{j=1, j \neq k}^{m} \frac{(n_j - 1 + g_j)!}{(n_j - 1)! g_j!} \left(\frac{-i}{r_j - r_k} \right)^{n_j + g_j} \right\}$$

$$= e^t \left[\prod_{l=1}^{m} r_l^{n_l} \right] \sum_{k=1}^{m} \left\{ \sum_{\substack{g_1, g_2, \ldots, g_m = 0 \\ \sum g_i = n_k - 1}} \frac{(t)^{g_k}}{g_k!} e^{-r_k t} \right.$$

$$\times \prod_{j=1, j \neq k}^{m} \frac{(n_j - 1 + g_j)!}{(n_j - 1)! g_j!} \frac{(-1)^{g_j}}{(r_j - r_k)^{n_j + g_j}} \right\} . \qquad (16)$$

Comparing eq. (16) with eqs. (12) and (13), we see that the right hand side of (16) is composed of the product of the factor e^t and $f_e(t)$ of eq. (13). In fact, the explicit expression for $f_e(t)$, in addition to the new derivation presented here in eq. (16), was derived

much earlier [27] under different context and was rediscovered/rederived multiple times [26,28,29] by different means. Its connection to combining P-values, however, was never made explicit until now.

From (10), we know that $F(\tau = 0) = 0$, implying that

$$
\begin{aligned}
F(\tau) &= \int_0^\tau f(\tau')d\tau' = \int_t^\infty f(e^{-t'})\, e^{-t'}dt' \\
&= \left[\prod_{l=1}^m r_l^{n_l} \right] \sum_{k=1}^m \sum_{\substack{g_1,g_2,\dots,g_m=0 \\ \sum g_i = n_k-1}} \left(\prod_{j=1,j\neq k}^m \frac{(n_j-1+g_j)!}{(n_j-1)!g_j!} \frac{(-1)^{g_j}}{(r_j-r_k)^{n_j+g_j}} \right) \times \\
&\quad \int_t^\infty \frac{(t')^{g_k}}{g_k!} e^{-r_k t'}dt' \\
&= \left[\prod_{l=1}^m r_l^{n_l} \right] \sum_{k=1}^m \sum_{\substack{g_1,g_2,\dots,g_m=0 \\ \sum g_i = n_k-1}} \left(\prod_{j=1,j\neq k}^m \frac{(n_j-1+g_j)!}{(n_j-1)!g_j!} \frac{(-1)^{g_j}}{(r_j-r_k)^{n_j+g_j}} \right) \times \\
&\quad \left(\sum_{l=0}^{g_k} \frac{t^{g_k-l}}{r_k^{l+1}(g_k-l)!} e^{-r_k t} \right) \\
&= \sum_{k=1}^m \sum_{\substack{g_1,g_2,\dots,g_m=0 \\ \sum g_i = n_k-1}} \left(\prod_{j=1,j\neq k}^m \frac{(n_j-1+g_j)!}{(n_j-1)!g_j!} \frac{(-r_k)^{g_j}r_j^{n_j}}{(r_j-r_k)^{n_j+g_j}} \right) \left(\sum_{l=0}^{g_k} \frac{(r_k t)^{g_k-l}}{(g_k-l)!} e^{-r_k t} \right) \\
&= \sum_{k=1}^m \sum_{g_k=0}^{n_k-1} \sum_{\substack{g_i\neq k=0 \\ \sum_i g_i = n_k-1}}^{n_k-1-g_k} \left(\prod_{j=1,j\neq k}^m \frac{(n_j-1+g_j)!}{(n_j-1)!g_j!} \frac{(-r_k)^{g_j}r_j^{n_j}}{(r_j-r_k)^{n_j+g_j}} \right) H(r_k t, g_k) ,
\end{aligned}
\tag{17}
$$

where the function H is defined as

$$
H(x,n) \equiv e^{-x} \sum_{k=0}^n \frac{x^k}{k!} .
\tag{18}
$$

Eq. (17) represents the most general formula that interpolates the scenarios considered by both Fisher and Good.

Let us take the limiting cases from (17). For Fisher's formula, one weights every P-value equally, and thus corresponds to $m=1$ and $n_1=L$. The constraint in the sum of (17) forces $g_1 = n_1 - 1 = L - 1$. Consequently, we have (by calling r_1 by r for simplicity)

$$
\text{Prob}(Q_F \leq \tau_F) = H(rt, L-1) = e^{-rt} \sum_{l=0}^{L-1} \frac{(rt)^l}{l!} .
\tag{19}
$$

Notice that regardless whatever the weight w one assigns to all the P-values, the final answer is independent of the weight. This is because $t = -\ln \tau = -w \ln \tau_F = (-\ln \tau_F)/r$ and therefore $rt = \ln(1/\tau_F)$. This results in

$$
\text{Prob}(Q_F \leq \tau_F) = \tau_F \sum_{l=0}^{L-1} \frac{[\ln(1/\tau_F)]^l}{l!} ,
\tag{20}
$$

exactly what one anticipates from (5). To obtain the results of Good, one simply makes $m=L$ and $n_k=1$ $\forall k$, implying all $g_i=0$. In this case, (17) becomes (with $r_l = 1/w_l$, $e^{-t} = \tau_G$ and $H(r_l t, 0) = e^{-r_l t} = \tau_G^{1/w_l}$)

$$
\text{Prob}(Q_G \leq \tau_G) = \sum_{l=1}^L \left(\prod_{k\neq l} \frac{r_k}{r_k - r_l} \right) \tau_G^{1/w_l} = \sum_{l=1}^L \Lambda_l \, \tau_G^{1/w_l} ,
\tag{21}
$$

reproducing exactly (6).

One may also re-express eq. (17) in a slightly different form

$$
\begin{aligned}
F(\tau) &= \left[\prod_{l=1}^m r_l^{n_l} \right] \sum_{k=1}^m \sum_{g_k=0}^{n_k-1} \frac{1}{r_k^{g_k+1}} H(r_k t, g_k) \times \\
&\quad \sum_{\substack{g_{i\neq k}=0 \\ \sum_i g_i = n_k-1}}^{n_k-1-g_k} \left(\prod_{j=1,j\neq k}^m \frac{(n_j-1+g_j)!}{(n_j-1)!g_j!} \frac{(-r_k)^{g_j}r_j^{n_j}}{(r_j-r_k)^{n_j+g_j}} \right) \\
&\equiv \left[\prod_{l=1}^m r_l^{n_l} \right] \tilde{F}(\tau; n_1, n_2, \dots, n_m) .
\end{aligned}
\tag{22}
$$

Note that in the expression (22), we have isolated an overall multiplying factor and have kept explicit the $n_{1\leq k\leq m}$ dependence for later convenience. As cautioned by Good [10] regarding Good's formula, the products of the inverse weight differences in eq. (22) may cause numerical instability in computing the combined P-values when some of the inverse weights become nearly degenerate. To see this point, let us consider varying r_j from a bit smaller than r_k to a bit larger than r_k. Although the change of weight $1/r_j$ is infinitesimal, some terms in (22) do change abruptly. We will provide some numerical examples in the Example subsection.

Accommodations of arbitrary weights

In our derivation of (20) in the previous subsection, it is explicitly shown that the final P-value obtained is independent of the weight w that was assigned to all the individual P-values, $p_1, p_2, \dots, p_L$. It is thus natural to ask, if one starts by weighing each P-value differently, upon making the weights close to one another, will one recover Fisher's formula (5) from Good's formula (6) in the limit of degenerate weights? By continuity, the answer is expected to be affirmative. In the broader context of the GC, one would like to have a formal protocol to compute the combined P-value when some of the weights become (nearly) degenerate.

In this subsection, we first illustrate the transition from Good's formula to Fisher's formula by combining two P-values with almost degenerate weights. We will then provide a general protocol to deal explicitly with the numerical instability caused by nearly degenerate weights. Possible occurrences of this instability were first cautioned by Good [10] and subsequently by many authors [13–15].

Let us consider combining p_1 and p_2 with weights w_1 and w_2 using Good's formula. One has

$$
\text{Prob}(Q_G \leq \tau_G) = \frac{1}{w_1 - w_2} \left[w_1 p_1 p_2^{\frac{w_2}{w_1}} - w_2 p_1^{\frac{w_1}{w_2}} p_2 \right] .
\tag{23}
$$

Without loss of generality, one assumes $w_1 > w_2$ and hence writes $w_1/w_2 = 1 + \epsilon$ with $\epsilon > 0$. We are interested in the case when the weights get close to each other, or when $\epsilon \to 0$. We now rewrite eq. (23) as

$$\mathrm{Prob}(Q_G \leq \tau_G) = \frac{w_2}{w_1 - w_2}\left[\frac{w_1}{w_2} p_1 p_2^{\frac{w_2}{w_1}} - p_1^{\frac{w_1}{w_2}} p_2\right]$$

$$= \frac{1}{\epsilon}\left[(1+\epsilon)p_1 p_2^{\frac{1}{1+\epsilon}} - p_1^{1+\epsilon} p_2\right] . \tag{24}$$

In the limit of small ϵ, we may rewrite (24) as

$$\mathrm{Prob}(Q_G \leq \tau_G) = \frac{p_1 p_2}{\epsilon}\left[(1+\epsilon)p_2^{-\frac{\epsilon}{1+\epsilon}} - p_1^{\epsilon}\right]$$

$$= \frac{p_1 p_2}{\epsilon}\left[(1+\epsilon)e^{-\frac{\epsilon}{1+\epsilon}\ln p_2} - e^{\epsilon \ln p_1}\right]$$

$$= \frac{p_1 p_2}{\epsilon}\left[\epsilon - \epsilon(\ln p_2 + \ln p_1) + O(\epsilon^2)\right] \tag{25}$$

$$= p_1 p_2 [1 - \ln(p_1 p_2) + O(\epsilon)]$$

$$\xrightarrow[\epsilon \to 0]{} p_1 p_2 [1 - \ln(p_1 p_2)] = \mathrm{Prob}(Q_F \leq \tau_F)$$

Note that when the small weight difference $w_1 - w_2$ is near the machine precision of a digital computer, using formula (6) directly will inevitably introduce numerical instability caused by rounding errors.

To construct a protocol to deal with nearly degenerate weights, one first observes from eqs. (14–22) that it is the inverse weights $r_k \equiv 1/w_k$ that permeate the derivation of the unified P-value. The closeness between weights is thus naturally defined by closeness in the inverse weights. As shown in eqs. (2) and (6), the combined P-value yielded by Good's formula depends only on the pairwise ratios of the weights. Making the observation that $r_k t$ in eq. (17) only depends on the ratios $r_k/r_{j \neq k}$, one deduces that for the GC the combined P-values (see (17)) also depend only on the ratios of weights, not the individual weights. We are thus free to choose any scale we wish. For simplicity, we normalize the inverse weight associated with each method by demanding the sum of inverse weights equal the total number of methods

$$\sum_{j=1}^{L} r_j = \sum_{j=1}^{L} 1 = L , \tag{26}$$

where $1/r_j$ represents the weight associated method j and L represents the total number of P-values (or methods) to be combined. For the GC described in the Methods section, $L = \sum_{k=1}^{m} n_k$. This normalization choice makes the average inverse weight of participating methods be 1.

The next step is to determine, for a given list of inverse weights and the radius for clustering, the number of clusters needed. This task may be achieved in a hierarchical manner. After normalizing the inverse weights r_k using eq. (26), one may sort the inverse weights in either ascending or descending order. For a given radius $\eta > 0$, one starts to seek the pair of inverse weights that are closest but not identical, and check if their difference is smaller than the radius η. If yes, one will merge that pair of inverse weights by using their average, weighted by number of occurrences, as the new center and continue the process until every inverse weight in the list is separated by a distance farther than η. We use an example of $L = 8$ to illustrate the idea. Let the normalized inverse weights $\{r_j\}_{j=1}^{8}$ be

$$0.50, 0.70(2), 0.71, 0.74, 1.03, 1.80, 1.82,$$

where the number 2 inside the pair of parentheses after 0.70 simply indicates that there are two identical inverse weights 0.70 to start with. Assume that one chooses η, the radius for clustering, to be 0.005. Since every pair of adjacent inverse weights are separated by more than 0.005, no further clustering procedures is needed and one ends up having seven effective clusters: one cluster with two identical inverse weights 0.70, and six singletons. This corresponds to $m = 7$, $n_1 = 1$, $n_2 = 2$, $n_3 = n_4 = \cdots = n_7 = 1$.

Suppose one chooses the clustering radius η to be 0.05. In the first step, we identify that 0.70 and 0.71 are the closest pair of inverse weights. The weighted average between them is

$$\frac{2 \cdot 0.70 + 0.71}{3} = \frac{2.11}{3} = 0.70\bar{3} .$$

The list of inverse weights then appears as

$$0.50, 0.70\bar{3}(3), 0.74, 1.03, 1.80, 1.82 .$$

The closest pair of inverse weights is now between 1.80 and 1.82, and upon merging them the list becomes

$$0.50, 0.70\bar{3}(3), 0.74, 1.03, 1.81(2) .$$

The next pair of closest inverse weights is then $0.70\bar{3}$ and 0.74. The weighted average leads to $(2.11 + 0.74)/4 = 0.7125$. After this step, the difference between any two cluster centers is larger than 0.05. The list of inverse weights now appears as

$$0.50, 0.7125(4), 1.03, 1.81(2) ,$$

indicating that we have $m = 4$ (four clusters), with number of members being $n_1 = 1$, $n_2 = 4$, $n_3 = 1$ and $n_4 = 2$. The centers of the four clusters are specified by the averaged inverse weights: $0.50, 0.7125, 1.03, 1.81$.

This is a good place for us to introduce some notation. We shall denote by $r_k + \eta_{k;j}$ the jth inverse weights of cluster k, whose averaged inverse weight is r_k. With this definition, for the example above, we have $\eta_{1;1} = 0$, $\eta_{2;1} = \eta_{2;2} = -0.0125$, $\eta_{2;3} = -0.0025$, $\eta_{2;4} = 0.0275$, $\eta_{3;1} = 0$, $\eta_{4;1} = -0.01$, and $\eta_{4;2} = 0.01$.

Using the hierarchical protocol mentioned above, the number of clusters m, the center r_k of the inverse weight of cluster k, and the numbers of members n_k of cluster k are all obtained along with $\{\eta_{k;j}\}$ once and for all. The $\eta_{k;j}$, as will be shown later, constitute the key expansion parameters that yield, upon multiplying by $\tilde{F}(\tau; n_l)$ with different $\{n_l\}$, the higher order terms in our key result. We show below how this is done.

Following the derivation in the previous subsection, we obtain a probability density function very similar to (14)

$$f(\tau) =$$

$$\left[\prod_{l=1}^{m}\prod_{j=1}^{n_l}(r_l + \eta_{l;j})\right](i)^{\sum_{k=1}^{m} n_k}\int_{-\infty}^{\infty}\frac{dq}{2\pi}e^{-it(q+i)}\prod_{k=1}^{m}\left[\prod_{j=1}^{n_k}\frac{1}{q+i(r_k+\eta_{k;j})}\right] . \tag{27}$$

From the preceding subsection, we see that the ill-conditioned situations emerge when some weights are nearly degenerate and the source of difference in inverse weights comes from obtaining $\tilde{F}(\tau; n_1, n_2, \ldots, n_m)$ in (22) from $\tilde{f}(\tau; n_1, n_2, \ldots, n_m)$ in (15). Therefore, one may leave the prefactor $\left[\Pi_{l=1}^{m} \Pi_{j=1}^{n_l} (r_l + \eta_{l;j}) \right]$ untouched and focus on the rest of the right hand side of eq. (27). To proceed, we write

$$\frac{1}{q + i(r_k + \eta_{k;j})} = \frac{1}{q + ir_k} \left(1 + \frac{i\eta_{k;j}}{q + ir_k} \right)^{-1}$$

$$= \frac{1}{q + ir_k} e^{-\ln\left(1 + \frac{i\eta_{k;j}}{q + ir_k}\right)} = \frac{1}{q + ir_k} \exp\left[\sum_{g=1}^{\infty} \frac{1}{g} \left(\frac{-i\eta_{k;j}}{q + ir_k} \right)^g \right].$$

Consequently, we may write

$$\prod_{j=1}^{n_k} \frac{1}{q + i(r_k + \eta_{k;j})} = \frac{1}{(q + ir_k)^{n_k}} \exp\left[\sum_{g=1}^{\infty} \frac{Y_{k;g}(i)^g}{(q + ir_k)^g} \right], \qquad (28)$$

where

$$Y_{k;g} \equiv \sum_{j=1}^{n_k} \frac{(-\eta_{k;j})^g}{g} . \qquad (29)$$

The product in eq. (27) may now be formally written as

$$\prod_{k=1}^{m} \left[\prod_{j=1}^{n_k} \frac{1}{q + i(r_k + \eta_{k;j})} \right]$$

$$= \left[\prod_{k=1}^{m} \frac{1}{(q + ir_k)^{n_k}} \right] \exp\left[\sum_{g=1}^{\infty} (i)^g \sum_{k=1}^{m} \frac{Y_{k;g}}{(q + ir_k)^g} \right]. \qquad (30)$$

We now note a simplification by choosing r_k to be the average inverse weight of the kth cluster. In this case, we have $\sum_{j=1}^{n_k} \eta_{k;j} = 0 \ \forall k$. That is, $Y_{k;1} = 0$ always. This allows us to write eq. (30) as

$$\prod_{k=1}^{m} \left[\prod_{j=1}^{n_k} \frac{1}{q + i(r_k + \eta_{k;j})} \right]$$

$$= \left[\prod_{k=1}^{m} \frac{1}{(q + ir_k)^{n_k}} \right] \exp\left[\sum_{g=2}^{\infty} (i)^g \sum_{k=1}^{m} \frac{Y_{k;g}}{(q + ir_k)^g} \right]. \qquad (31)$$

The key idea here is to Taylor expand the exponential and collect terms of equal number of $1/(q + ir)$. Evidently, the first correction term starts with $1/(q + ir)^2$. Furthermore, before the $1/(q + ir)^4$ order, there is no mixing between different clusters. Below, we rewrite eq. (27) to include the first few orders of correction terms

$$\frac{f(\tau)}{\Pi_{l=1}^{m} \Pi_{j=1}^{n_l} (r_l + \eta_{l;j})}$$

$$= (i)^{\sum_{k=1}^{m} n_k} \int_{-\infty}^{\infty} \frac{dq}{2\pi} e^{-it(q+i)} \frac{\exp\left[\sum_{g=2}^{\infty} (i)^g \sum_{k=1}^{m} \frac{Y_{k;g}}{(q + ir_k)^g} \right]}{\Pi_{k=1}^{m} (q + ir_k)^{n_k}}$$

$$= \tilde{f}(\tau; \{n_l\}_{l=1}^{m}) + \sum_{k=1}^{m} Y_{k;2} \, \tilde{f}(\tau; \{n_{l \neq k}, n_k + 2\})$$

$$+ \sum_{k=1}^{m} Y_{k;3} \, \tilde{f}(\tau; \{n_{l \neq k}, n_k + 3\})$$

$$+ \sum_{k=1}^{m} \left(Y_{k;4} + \frac{(Y_{k;2})^2}{2!} \right) \tilde{f}(\tau; \{n_{l \neq k}, n_k + 4\})$$

$$+ \frac{1}{2!} \sum_{\substack{k,k'=1 \\ k \neq k'}}^{m} Y_{k;2} Y_{k';2} \, \tilde{f}(\tau; \{n_{l \neq k,k'}, n_k + 2, n_{k'} + 2\}) + O(\eta^5) . \qquad (32)$$

This immediately leads to

$$\frac{F(\tau)}{\Pi_{l=1}^{m} \Pi_{j=1}^{n_l} (r_l + \eta_{l;j})}$$

$$= \tilde{F}(\tau; \{n_l\}_{l=1}^{m}) + \sum_{k=1}^{m} Y_{k;2} \, \tilde{F}(\tau; \{n_{l \neq k}, n_k + 2\})$$

$$+ \sum_{k=1}^{m} Y_{k;3} \, \tilde{F}(\tau; \{n_{l \neq k}, n_k + 3\})$$

$$+ \sum_{k=1}^{m} \left(Y_{k;4} + \frac{(Y_{k;2})^2}{2!} \right) \tilde{F}(\tau; \{n_{l \neq k}, n_k + 4\})$$

$$+ \frac{1}{2!} \sum_{\substack{k,k'=1 \\ k \neq k'}}^{m} Y_{k;2} Y_{k';2} \, \tilde{F}(\tau; \{n_{l \neq k,k'}, n_k + 2, n_{k'} + 2\}) + O(\eta^5) \qquad (33)$$

Note that when the clustering radius η is chosen to be zero, the only clusters are from groups of *identical* weights, and all $\eta_{k;j}$ must be zero. In this case, only the first term on the right hand side of (33) exists and the result derived in the previous subsection is recovered exactly. Since all $\tilde{F}$ are finite positive quantities, the errors resulting from truncating the expression in eq. (33) at certain order of η can be easily bounded. Therefore, any desired precision may be obtained via including the corresponding number of higher order terms. As the main result of the current paper, our expansion provides a systematic, numerically stable method to achieve desired accuracy in computing combined *P*-values.

Examples

Example (a). This example provides a numerical work flow to compute the $\tilde{F}(\tau; \{n_l\})$ function present in eq. (22). Assuming $m = 4$, we show below how to open up the sum in eq. (22). The constraint $\sum_i g_i = n_k - 1$ implies that one only has $m - 1$ ($= 3$ here) independent g_is. Once the $(m-1)$ g_is are specified, the remaining one is determined. To simplify the exposition, let us

introduce the following notation

$$\alpha(g_j; j, k) \equiv \frac{(n_j - 1 + g_j)!}{(n_j - 1)! g_j!} \frac{(-1)^{g_j}}{(r_j - r_k)^{n_j + g_j}} \ .$$

This allows one to expand the sum in (22) as

$$
\begin{aligned}
\tilde{F}(\tau) =& \sum_{g_1=0}^{n_1-1} \frac{H(r_1 t, g_1)}{r_1^{g_1+1}} \sum_{g_2=0,}^{n_1-1-g_1} \alpha(g_2; 2, 1) \sum_{g_3=0}^{n_1-1-g_1-g_2} \alpha(g_3; 3, 1) \ \alpha(g_4; 4, 1) \\
+& \sum_{g_2=0}^{n_2-1} \frac{H(r_2 t, g_2)}{r_2^{g_2+1}} \sum_{g_1=0}^{n_2-1-g_2} \alpha(g_1; 1, 2) \sum_{g_3=0}^{n_2-1-g_2-g_1} \alpha(g_3; 3, 2) \ \alpha(g_4; 4, 2) \\
+& \sum_{g_3=0}^{n_3-1} \frac{H(r_3 t, g_3)}{r_3^{g_3+1}} \sum_{g_1=0}^{n_3-1-g_3} \alpha(g_1; 1, 3) \sum_{g_2=0}^{n_3-1-g_3-g_1} \alpha(g_2; 2, 3) \ \alpha(g_4; 4, 3) \\
+& \sum_{g_4=0}^{n_4-1} \frac{H(r_4 t, g_4)}{r_4^{g_4+1}} \sum_{g_1=0}^{n_4-1-g_4} \alpha(g_1; 1, 4) \sum_{g_2=0}^{n_4-1-g_4-g_1} \alpha(g_2; 2, 4) \ \alpha(g_3; 3, 4) \ .
\end{aligned}
\tag{34}
$$

Note that in eq. (33), in the zeroth order term, the argument n_l of $\tilde{F}$ represents the number of members associated with cluster l. However, for higher order correction terms, the n_ls entering $\tilde{F}$ no longer carry the same meaning. Therefore, in the example shown here, one should not assume that n_j is the number of methods associated with cluster j.

Example (b). This example illustrates the possibility of numerical instability associated with eqs. (6) and (22) when they are used to combine P-values with nearly equal weights. This instability arises from adding numbers with nearly identical magnitude but different signs, yielding a value containing few or no significant figures. We also show how such instabilities are resolved by using eq. (33). Consider the case of combining five P-values, {0.008000257, 0.008579261, 0.0008911761, 0.006967988, 0.004973110}, weighted respectively by {0.54531152, 0.54532057, 0.54531221, 0.54531399, 0.54531776}. Using eq. (2), one obtains $\tau_G = 4.30656196 \times 10^{-7}$. The combined P-value is then obtained as the probability of attaining a random variable Q_G, defined in eq. (4), such that it is less than or equal to τ_G.

Combining P-values using eq. (6) gives

$$
\begin{aligned}
\text{Prob}(Q_G \leq \tau_G) =& 1923475672.53812003 + 134195847.49348195 \\
& - 3271698577.16100168 + 1726093852.57087326 \\
& - 512066795.44147670 = -0.00000322 \ .
\end{aligned}
$$

When one uses equation (22), τ takes the value of τ_G and the random variable Q is simply Q_G, and the combined P-value becomes

$$
\begin{aligned}
\text{Prob}(Q \leq \tau) =& 170090507.09336647 + 21761086.68190728 \\
& - 972903041.25101399 + 941269625.31004059 \\
& - 512066795.44252247 = -0.00000006 \ .
\end{aligned}
$$

Apparently, probability cannot be negative. The negative values shown above illustrate how eqs. (6) and (22) may lead to cancellation of numbers of comparable magnitude thus may yield meaningless values when the weights are nearly degenerate. This numerical instability is removed by applying equation (33), which combines weighted P-values using a controlled expansion and

yields, for this example,

$$
\begin{aligned}
\text{Prob}(Q \leq \tau) =& 5.379093 \times 10^{-8} + 1.407305 \times 10^{-16} \\
& - 1.066323 \times 10^{-21} + 1.634917 \times 10^{-25} + O(10^{-29}) \\
=& 5.37909 \times 10^{-8} \ .
\end{aligned}
$$

Example (c). One natural question to ask is how well does eq. (33) work when one chooses a larger clustering radius and group weights that are clearly distinguishable into a few clusters? To consider this case, let us use the five P-values from example (b) but with weights chosen differently. Let us assume that the inverse weights ($r_k \equiv 1/w_k$) associated with these five P-values are {0.6, 0.65, 1.2, 1.25, 1.3}. For this case, $\tau = \tau_G = 1.935663 \times 10^{-13}$. Combining P-value using formulas (6) yields

$$
\begin{aligned}
\text{Prob}(Q \leq \tau) =& 2.187324 \times 10^{-6} - 5.946040 \times 10^{-7} \\
& + 2.131226 \times 10^{-13} - 8.011644 \times 10^{-14} \\
& + 7.639290 \times 10^{-15} = 1.59272 \times 10^{-6} \ ,
\end{aligned}
$$

while combining P-values using (22) yields identical results

$$
\begin{aligned}
\text{Prob}(Q \leq \tau) =& 1.725699 \times 10^{-6} - 3.049251 \times 10^{-7} \\
& + 1.311524 \times 10^{-13} - 6.162803 \times 10^{-14} \\
& + 7.639290 \times 10^{-15} = 1.59272 \times 10^{-6} \ .
\end{aligned}
$$

When one uses $\eta = 0.1$ as the clustering radius, one obtains two clusters: one with average inverse weight 0.625 and the other with average inverse weight 1.25. If one then uses eq. (33) to combine P-values, one attains the following results

$$
\begin{aligned}
\text{Prob}(Q \leq \tau) =& \ 1.472453 \times 10^{-6} + 1.171521 \times 10^{-7} + 0 \\
& + 2.584710 \times 10^{-9} + 4.889899 \times 10^{-10} + O(10^{-12}) \\
=& 1.59268 \times 10^{-6} \ ,
\end{aligned}
\tag{35}
$$

which contains no sign alternation and agrees well with the results from both (6) and (22). This illustrates the robustness of eq. (33) in combining P-values. Note that the third term on the right hand side of (35) is zero. This is because the multiplying factor $Y_{k;3}$ is zero for both clusters. In general, $Y_{k;3}$ measures the skewness of inverse weights associated with cluster k and for our case here both clusters of inverse weights are perfectly symmetrical with respect to their centers, leading to zero skewness. If the inverse weights of cluster k distribute perfectly symmetrically with respect to its center, it is evident from eq. (29) that $Y_{k;g} = 0$ for odd g.

Evidently if one chooses a large clustering radius η and then uses eq. (33) to combine P-values, many higher order terms in the expansion will be required to achieve high accuracy in the final combined P-value.

Discussion

Although the expression (17) provides access to exact statistics for a broader domain of problems and our expansion formula (33) provides accurate and stable statistics even when nearly degenerate weights are present, there remain a few unanswered questions

that should be addressed by the community in the near future. For example, even though we can accommodate any reasonable P-value weighting, thanks to (33), the more difficult question is how does one choose the right set of weights when combining statistical significance [35–39]. The weights chosen should reflect how much one wishes to trust various obtained P-values. Ideally, a fully systematic method should also provide a metric for choosing appropriate weights. How to obtain the best set of weights remains an open problem and definitely deserves further investigations.

Another limitation of (17) and (33), and consequently of Fisher's and Good's formulas, is that one must assume the P-values to be combined are independent. In real applications, it is foreseeable that P-values reported by various methods may exhibit non-negligible correlations. How to obtain the correlation [40–42] and how to properly incorporate P-value correlations [15,43,44] while combining P-values are challenging problems that we hope to address in the near future.

Author Contributions

Conceived and designed the experiments: YKY. Performed the experiments: GA YKY. Analyzed the data: GA YKY. Wrote the paper: YKY.

References

1. Stouffer S, Suchman E, DeVinney L, Star S, Williams RMJ (1949) The American Soldier, Vol. 1: Adjustment During Army Life. Princeton: Princeton University Press.
2. Mosteller F, Bush RR (1954) Selected Quantitative Techniques. In: Handbook of Social Psychology, volume 1. Cambridge, Mass.: Addison-Wesley. pp 289–334.
3. Olkin I (1995) Statistical and theoretical considerations in meta-analysis. Journal of Clinical Epidemiology 48: 133–146.
4. Loesgen S, Dempfle A, Golla A, Bickeboller H (2001) Weighting schemes in pooled linkage analysis. Genet Epidemiol 21 Suppl 1: S142–7.
5. Alves G, Wu WW, Wang G, Shen RF, Yu YK (2008) Enhancing Peptide Identification Confidence by Combining Search Methods. J Proteome Res 7: 3102–3113.
6. Hess A, Iyer H (2007) Fisher's combined p-value for detecting differentially expressed genes using Affymetrix expression arrays. BMC Genomics 8: 96.
7. Bailey TL, Gribskov M (1998) Combining Evidence using P-values: Application to Sequence Homology Searches. Bioinformatics 14: 48–54.
8. Yu YK, Gertz EM, Agarwala R, Schaffer AA, Altschul SF (2006) Retrieval accuracy, statistical significance and compositional similarity in protein sequence database searches. Nucleic Acids Res 34: 5966–73.
9. Fisher RA (1932) Statistical Methods for Research Workers, Vol II. Edinburgh: Oliver and Boyd.
10. Good IJ (1955) On the weighted combination of significance tests. Journal of the Royal Statistical Society Series B (Methodological) 17: 264–265.
11. Solomon H, Stephens MA (1977) Distribution of a sum of weighted chi-square variables. Journal of the American Statistical Association 72: 881–885.
12. Gabler S (1987) A quick and easy approximation to the distribution of a sum of weighted chi-square variables. Statistische Hefte 28: 317–25.
13. Bhoj DS (1992) On the distribution of the weighted combination of independent probabilities. Statistics & Probability Letters 15: 37–40.
14. Olkin I, Saner H (2001) Approximations for trimmed Fisher procedures in research synthesis. Stat Methods Med Res 10: 267–76.
15. Hou CD (2005) A simple approximation for the distribution of the weighted combination of non-independent or independent probabilities. Statistics & Probability Letters 73: 179–187.
16. Tippett L (1931) The Methods of Statistics. London: Williams and Norgate Ltd.
17. Pearson K (1933) On a method of determining whether a sample of size n supposed to have been drawn from a parent population having a known probability integral has probably been drawn at random. Biometrika 25: 379–410.
18. Pearson ES (1938) The probability integral transformation for testing goodness of fit and combining independent tests of significance. Biometrika 30: 134–148.
19. Bateman H (1910) A solution of a system of differential equations occurring in the theory of radioactive transformation. Pro Cambridge Philosophical Soc 15: 423–427.
20. Bahrucha-Reid A (1960) Elements of the Theory of Markov Processes and Their Applications McGraw-Hill.
21. McGill WJ, Gibbon J (1965) The general-gamma distribution and reaction times. Journal of Mathematical Psychology 2: 1–18.
22. Likes J (1967) Distributions of some statistics in samples from exponential and powerfunction populations. Journal of the American Statistical Association 62: 259–271.
23. Morse P (1958) Queues, Inventories and Maintenance. John Wiley & Sons.
24. Cox D (1962) Renewal Theory. Methuen & Co.
25. MacNeill IB (1974) Tests for change of parameter at unknown times and distributions of some related functionals on brownian motion. The Annals of Statistics 2: 950–962.
26. Jasiulewicz H, Kordecki W (2003) Convolutions of Erlang and of Pascal Distributions with Applications to Reliability. Demonstratio Mathematica 36: 231–238.
27. Mathai A (1983) On linear combinations of independent exponential variables. Communications in Statistics - Theory and Methods 12: 625–632.
28. Harrison PG (1990) Laplace transform inversion and passage-time distribution in markov processes. J Appl Prob 27: 74–87.
29. Amari S, Mirsa RB (1997) Closed-form Expression for Distribution of the Sum of Independent Exponential Random Variables. IEEE Trans Reliability 46: 519–522.
30. Whitlock MC (2005) Combining probability from independent tests: the weighted Zmethod is superior to Fisher's approach. J Evol Biol 18: 1368–73.
31. Liptak P (1958) On the Combination of Independent Tests. Magyar Tud Akad Nat Kutato int Kozl 3: 171–197.
32. Koziol JA, Tuckwell HC (1994) A Weighted Nonparametric Procedure for the Combination of Independent Events. Biom J 36: 1005–1012.
33. Lancaster HD (1961) The combination of probabilities: An application of orthogonal functions. Austr J Statist 3: 20–33.
34. Koziol JA (1996) A Note on Lancaster's Procedure for the Combination of Independent Events. Biometrical Journal 38: 653–660.
35. Zelen M, Joel LS (1959) The weighted compounding of two independent significance tests. The Annals of Mathematical Statistics 30: 885–895.
36. Koziol J, Perlman M (1978) Combining Independent Chi-squared Tests. J Amer Statist Assoc 73: 753–763.
37. Hedges LV, Olkin I (1985) Statistical Methods for Meta-Analysis. Academic Press.
38. Pepe MS, Fleming TR (1989) Weighted Kaplan-Meier statistics: a class of distance tests for censored survival data. Biometrics 45: 497–507.
39. Forrest WF (2001) Weighting improves the "new Haseman-Elston" method. Hum Hered 52: 47–54.
40. Wei LJ, Lachin JM (1984) Two-sample asymptotically distribution-free tests for incomplete multivariate observations. Journal of the American Statistical Association 79: 653–661.
41. Pocock SJ, Geller NL, Tsiatis AA (1987) The analysis of multiple endpoints in clinical trials. Biometrics 43: 487–498.
42. James S (1991) Approximate multinormal probabilities applied to correlated multiple endpoints in clinical trials. Stat Med 10: 1123–35.
43. Brown MB (1975) A method for combining non-independent, one-sided tests of significance. Biometrics 31: 987–992.
44. Kost JT, McDermott MP (2002) Combining dependent p-values. Statistics & Probability Letters 60: 183–190.