

Methods to estimate the variance of some indices of the signal detection theory: A simulation study

Manuel Suero^{*1}, Jesús Privado², and Juan Botella¹

¹ *Universidad Autónoma de Madrid*

² *Universidad Complutense de Madrid*

A simulation study is presented to evaluate and compare three methods to estimate the variance of the estimates of the parameters δ and C of the signal detection theory (SDT). Several methods have been proposed to calculate the variance of their estimators, d' and c . Those methods have been mostly assessed by comparing the empirical means and variances in simulation studies with the calculations done with the parametric values of the probabilities of giving a yes response on a signal trial (hits) and on a noise trial (false alarms). In practical contexts the variance must be estimated from estimations of those probabilities (empirical rates of hits and false alarms). The three methods to estimate the variance compared in the present simulation study are based in the binomial distribution of Miller, the normal approach of Gourevitch and Galanter and the maximum likelihood method proposed by Dorfman and Alf. They are compared in terms of relative bias (accuracy) and the mean squared error (precision). The results show that the last two methods behave indistinguishably for practical purposes and provide severe over-estimation errors in a range of situations that while not the most common are perfectly credible in several practical contexts. By contrast, the method of Miller provides better results (or at least similar) in all conditions studied. It is the recommended method to obtain estimates of the variances of these statistics for practical purposes.

* Financial support: Project “Meta-análisis con índices de la Teoría de la Detección de Señales”. (Reference: PSI2013-45513; MINECO). Corresponding author: Manuel Suero. Facultad de Psicología. Universidad Autónoma de Madrid. c/ Ivan Pavlov, 6. 28049 Madrid, SPAIN. Phone: 34-914973241. E-mail: manuel.suero@uam.es

Sometimes we are interested in knowing the variance of the estimates of the parameters of the signal detection theory (SDT; Green & Swets, 1966; MacMillan & Creelman, 2005; Wickens, 2001). For example, for testing hypotheses about those parameters (Jesteadt, 2005; Miller, 1996; Verde, MacMillan, & Rotello, 2006) or when we want to perform a meta-analytic integration of the evidence on a specific issue and the studies have been conducted analyzing statistics associated with the SDT. Specifically, the indices d' and c , estimators of δ and C , which are measures of sensitivity and response bias respectively, are often used to reflect the effects sought in experimental studies (Logan, 2004; Swets, Dawes, & Monahan, 2000). In a meta-analysis the statistics provided by the primary studies are combined to yield a point estimate of the effect size. The most common method to combine the estimates consists in calculating a weighted average of the values, using as weights the reciprocals of their variances (weight = $1/\sigma^2$) (Borenstein, Hedges, Higgins, & Rothstein, 2009; Botella & Sánchez-Meca, 2015; Hedges & Olkin, 1985). Accordingly, to implement such a procedure it is necessary to know the variance of these statistics in each primary study.

Assuming the normal homoscedastic (NH) SDT model, and the yes/no experimental paradigms (MacMillan & Creelman, 2005), three main methods have been proposed for calculating the variance of d' (see below for a detailed technical presentation): the exact method of Miller (1996), the approximate method of Gourevitch and Galanter (1967), and the maximum likelihood method of Dorfman and Alf (1968).

The methods of Miller (1996) and Gourevitch and Galanter (1967) compute the variance of d' substituting in their formulas the conditional probabilities of a false alarm and a hit. The variance of d' is properly calculated when using those probabilities, and the obtained value is the true (parametric) variance. But this only can be done if the true probabilities are known, as in simulation studies. In those contexts where the true probabilities are unknown the variance of d' is calculated using the proportions of hits and false alarms obtained in a finite number of trials. Consequently, because these are estimators of the probabilities, the variance is an estimator of the true (parametric) variance.

Two studies have been conducted in order to compare some of the three methods. Miller (1996) calculated the variance of d' applying his procedure and the method of Gourevitch and Galanter. The variance of d' was calculated for different values of δ and number of trials with the same response bias (unbiased responding). The results show that: a) the variance of d' calculated by Miller's method and δ is increasing: the variance of d' increases to a maximum and then decreases, where the position of this

maximum (a δ value) depends on the number of trials; b) the variance of d' calculated by the method of of Gourevitch and Galanter and δ is increasing: the variance of d' increases.

In a Monte Carlo study, Kadlec (1999) compares the empirical variance of d' obtained in the simulation with those calculated using the method of Gourevitch and Galanter. Three variables were manipulated: δ , the number of trials and response bias. According to the results of Kadlec (Figure 10), the variance obtained by the method of Gourevitch and Galanter is similar to the empirical variance until a critical δ . Above this critical value, the method of Gourevitch and Galanter overestimated the variance of d' . The critical δ value depends on the number of trials and the response bias.

It is important to mention that in Miller (1996) the variance of d' was calculated using the parametric probabilities of false alarms and hits; on the contrary, in Kadlec (1999) the variance of d' following Gourevitch and Galanter was calculated using the proportions of false alarms and hits. Therefore, in the work of Miller (1996) the parametric value of the variance is calculated, whereas in Kadlec's study (1999) the estimators were obtained.

The difference referred to in the preceding paragraph makes it difficult to establish common conclusions of the two studies. Moreover, since in most practical situations the values of the probabilities of false alarms and hits are unknown, the estimation of the variance of d' must be calculated using proportions of false alarms and hits. Therefore, when the methods are compared, it is more useful to make these comparisons by means of the estimator of the variance.

In this paper we assess through simulation the suitability of three proposed methods to estimate the variance of d' and c in yes/no experimental paradigms (MacMillan & Creelman, 2005): method of Miller (1996), method of Gourevitch and Galanter (1967), and method of Dorfman and Alf (1968). Note that in the two above mentioned studies has not been evaluated the method of Dorfman and Alf (1968). Our simulation provides an empirical estimate of the variance of d' , and estimates obtained by the three procedures. Furthermore, the estimates of the variance of d' will be compared with the parametric value of variance of d' calculated using the procedure of Miller (1996). The merits of the three methods are assessed by an evaluation of their bias and precision for a range of values of δ , C , and N . Note that in the two above-mentioned studies has not been evaluated the precision. In the study presented here it is possible to evaluate the accuracy as both the estimated variance of d' and its parametric value will be

calculated. We begin providing a brief sketch of the SDT indices, the two main methods proposed to calculate the variance of d' , and the three procedures proposed to estimate that variance. Then we describe the simulation and finally assess the results of the study reaching some conclusions and suggesting practical guidelines.

Signal Detection Theory (SDT) indices

There are many indices to characterize the performance in various contexts that can be analyzed from the SDT (MacMillan & Creelman, 2005). Although we have a variety of parametric indices that assume different assumptions, and a number of nonparametric indices, we focus here on the two parametric indices more widely employed. The first, δ , is a measure of sensitivity and is defined as the distance between the expected values of the variable of evidence for a target (signal) stimulus and a non-target (noise) stimulus, expressed in standard deviations metric. The second, C , is an index of the response criterion or response bias, which is defined as the distance between the reference value to choose the response and the value corresponding to the intersection between the distributions. Under the NH model the value of the intersection is equidistant from the expected values (figure 1). Put another way, it is assumed for the noise stimuli an approximate $N(0; 1)$ distribution and for the signal stimuli an approximate distribution $N(\delta; 1)$. Therefore, the sensitivity parameter, δ , is the mean of the signal distribution.

Suppose an experiment in which there are N_s trials with a signal and N_n trials with noise, and the answers contain H hits and F false alarms. We get the hits ratio, $P_H = H / N_s$, and the false alarms ratio, $P_F = F / N_n$. The estimates of sensitivity, d' , and the response criterion, c , can be calculated from P_H and P_F (Macmillan & Creelman, 2005). The d' statistic is defined as,

$$d' = \hat{z}_H - \hat{z}_F \quad [1]$$

where \hat{z}_H and \hat{z}_F are estimates of the values of the standard normal whose cumulate probabilities equal the probabilities to give a yes response to a target stimulus and to a noise stimulus, respectively. The corresponding empirical proportions of hits and false alarms, P_H and P_F , are estimations of the true probabilities, π_H and π_F , as these are unknown. That is, $\hat{z}_H = \Phi^{-1}(P_H)$ and $\hat{z}_F = \Phi^{-1}(P_F)$. Likewise, c is defined as,

$$c = -\frac{1}{2}(\hat{z}_H + \hat{z}_F) \quad [2]$$

In the example of figure 1 the distance between the expected values equals 2 standard deviations ($\delta = 2$) and the curves intersect at $z = 1$. The value corresponding to the response criterion stands at 0.5: half standard deviations to the left of the crossing value ($C = -0.5$). Consequently, when a target stimulus is presented the probability of a yes response, π_H , is 0.9332, whereas the probability of a yes response to a noise stimulus, π_F , is 0.3085.

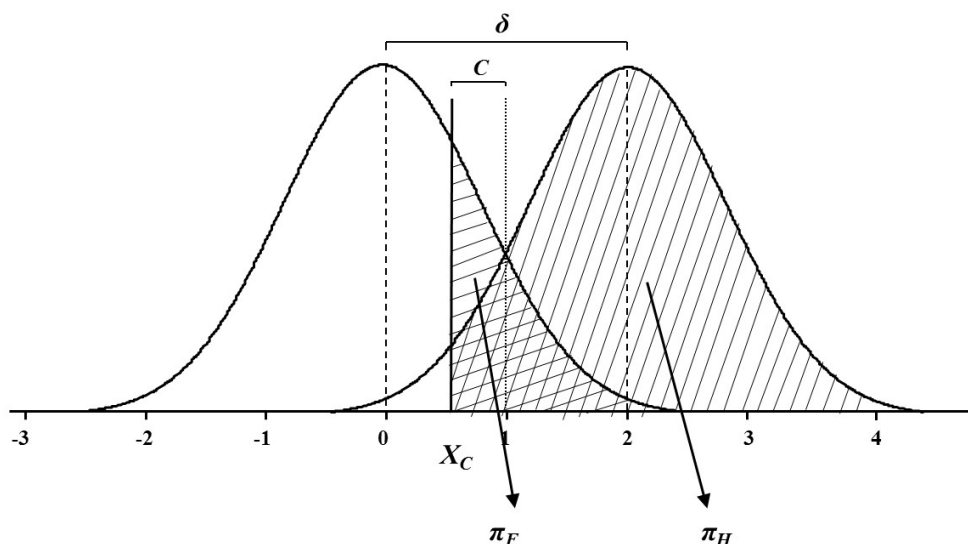


Figure 1. Example of the δ and C values for a specific case (see the text): $\pi_H = 0.9332$ and $\pi_F = 0.3085$.

Procedures to calculate the variance of d'

What indeed we are interested in are the parameters (δ and C), but what we know in virtually all practical occasions are their estimators, d' and c . When the information is collected through a limited number of trials with signal and noise (N_s and N_n), the statistics do not exactly match the parameter values, but show some deviation due to *sampling variance*. Knowing the sampling variance allows assess the properties of an estimator. We can choose among several alternative estimators the one having more suitable properties, based on its expected value and its variance. Specifically, other properties being equal it is preferable an unbiased

estimator; one for which the expected value is the own parameter intended to estimate. Furthermore, other properties being equal it is also preferable an estimator with high precision (low variance) because in the long run its values tend to look more to the parameter value.

Obtaining the variance of an estimator is not always easy or straightforward, as it might seem. In fact, as several proposals to calculate the variance of d' have been made, it is desirable to know which one (and in what conditions) provides estimates closer to the actual variance. We focus in d' because c is so closely related to it that the results and conclusions for d' can be safely generalized to c (see equations [1] and [2]).

Method of Gourevitch and Galanter

One of the first attempts to develop procedures to test hypotheses about δ and C is due to Gourevitch and Galanter (1967). They proposed an approach to the variance of d' assuming the NH model. Such an approach is obtained by developing a Taylor series of the standard normal distribution, considering only the first two terms of the series. With this procedure the following formula is reached by linear approximation,

$$\sigma_{d'}^2 \approx \frac{\pi_H \cdot (1 - \pi_H)}{N_s \cdot \phi^2(z_H)} + \frac{\pi_F \cdot (1 - \pi_F)}{N_n \cdot \phi^2(z_F)} \quad [3]$$

where π_H and π_F are, respectively, the probabilities of a hit and a false alarm; z_H and z_F are the values of the standard normal distribution associated, respectively, with cumulative probabilities equal to π_H and π_F ; N_s and N_n are the number of trials containing a signal and noise, respectively; and ϕ is the probability density function of the standard normal distribution.

Method of Miller

This author proposes a method for calculating the variance based on the exact distribution of \hat{z}_H and \hat{z}_F . Those two values are random variables distributed as binomials: $B(N_s; \pi_H)$ and $B(N_n; \pi_F)$. Taking in account [1] and that they are assumed independent:

$$\sigma_{d'}^2 = \sigma_{\hat{z}_H}^2 + \sigma_{\hat{z}_F}^2 \quad [4]$$

Following Miller (1996), the variance of \hat{z}_H is:

$$\sigma_{\hat{z}_H}^2 = \sum_{i=0}^{N_s} \left[\Phi^{-1}\left(\frac{i}{N_s}\right) \right]^2 \cdot \binom{N_s}{i} \cdot \pi_H^i \cdot (1 - \pi_H)^{N_s-i} - [E(\hat{z}_H)]^2 \quad [5]$$

where $E(\hat{z}_H)$ is the expected value of \hat{z}_H calculated by the expression:

$$E(\hat{z}_H) = \sum_{i=0}^{N_s} \Phi^{-1}\left(\frac{i}{N_s}\right) \cdot \binom{N_s}{i} \cdot \pi_H^i \cdot (1 - \pi_H)^{N_s-i} \quad [6]$$

Both in [5] and [6], Φ^{-1} is the inverse of the cumulative probability of the standard normal. Recalling that $\hat{z}_H \sim B(N_s; \pi_H)$ it is obvious that the expression [5] is the variance of the random variable \hat{z}_H , since

$$\hat{z}_H = \Phi^{-1}\left(\frac{i}{N_s}\right) \text{ and } \frac{i}{N_s} \text{ is } P_H.$$

The equation defining $\sigma_{\hat{z}_F}^2$ is similar to [5], but replacing N_s by N_n and π_H by π_F . Likewise, the expected value of \hat{z}_F , $E(\hat{z}_F)$, would be obtained similarly to [6].

Procedures to obtain an estimation of the variance, $\hat{\sigma}_{d'}^2$

The problem with the two methods above is that to calculate the variance of the sensitivity statistic, $\sigma_{d'}^2$, with the formulae proposed by Gourevitch and Galanter (1967) and Miller (1996) it is necessary to know both π_H and π_F . But as in most practical contexts these values are unknown, their estimates must be used: P_H as an estimate of π_H and P_F as an estimate of π_F . Then, what can be obtained are estimates of the variance, $\hat{\sigma}_{d'}^2$. The variance estimated for the d' values is also a random variable, as it is calculated using the values of the variables P_H and P_F (as in the formulae of this section) instead of the constants π_H and π_F (as in [3], [5] and [6]).

Three methods to estimate that variance will be evaluated in the simulation study presented below: the two methods already described but using the sample estimates instead of the parametric probabilities, and a maximum likelihood method (Dorfman & Alf, 1968; Kaplan, 2009).

Method of Gourevitch and Galanter

The estimation method based on Gourevitch and Galanter (1967) replaces π_H with P_H and π_F with P_F in equation [3]; it reads as:

$$\hat{\sigma}_{d'}^2 \approx \frac{P_H \cdot (1 - P_H)}{N_s \cdot \varphi^2(\hat{z}_H)} + \frac{P_F \cdot (1 - P_F)}{N_n \cdot \varphi^2(\hat{z}_F)} \quad [7]$$

Method of Miller

Similarly, in the procedure of estimation based on Miller's (1996) method π_H and π_F are replaced with P_H and P_F in equations [5] and [6] as:

$$\hat{\sigma}_{\hat{z}_H}^2 = \sum_{i=0}^{N_S} \left[\Phi^{-1}\left(\frac{i}{N_S}\right) \right]^2 \cdot \binom{N_S}{i} \cdot P_H^i \cdot (1-P_H)^{N_S-i} - \left[\hat{E}(\hat{z}_H) \right]^2 \quad [8]$$

$$\hat{E}(\hat{z}_H) = \sum_{i=0}^{N_S} \Phi^{-1}\left(\frac{i}{N_S}\right) \cdot \binom{N_S}{i} \cdot P_H^i \cdot (1-P_H)^{N_S-i} \quad [9]$$

The equation defining $\hat{\sigma}_{z_F}^2$ is similar to [8], but replacing N_S by N_n and P_H by P_F . The same logic is applied to obtain the expected value, $\hat{E}(\hat{z}_F)$.

Method of Dorfman and Alf

The aim of the procedure proposed by these authors is to estimate the parameters involved. Unlike the previous two methods, instead of using the equations [1] and [2] for calculating the estimators they obtain the estimates d' and c using the method of maximum likelihood. Adapting the logarithm of the likelihood function (equation 4 in Dorfman and Alf, 1968) for the NH model and keeping C constant along the trials, this function is equal to:

$$\begin{aligned} \text{Log}L = & F \cdot \log \left[\Phi \left(-C - \frac{\delta}{2} \right) \right] + (N_n - F) \cdot \log \left[1 - \Phi \left(-C - \frac{\delta}{2} \right) \right] + \\ & H \cdot \log \left[\Phi \left(\frac{\delta}{2} - C \right) \right] + (N_s - H) \cdot \log \left[1 - \Phi \left(\frac{\delta}{2} - C \right) \right] \end{aligned} \quad [10]$$

To estimate the parameters δ and C they must be obtained the values that maximize the expression [10]. In addition, it is obtained the variance-covariance matrix of the estimators. In the main diagonal of this matrix can be found $\hat{\sigma}_{d'}^2$, as an estimate of $\sigma_{d'}^2$. Both the estimates and the variance are obtained by numerical methods, as for example *RSCORE* (Dorfman, 1982) or *ROCFIT* (Metz, 1989).

The problem with extreme frequencies

In order to apply most of the equations above sometimes is necessary to obtain z values associated with P_H and/or P_F ratios equal to 1 or 0. In those cases the corresponding z values are $\pm\infty$, respectively, and d' is undefined. So, $\hat{\sigma}_z^2$ cannot be calculated. Several alternatives have been proposed to face this problem (see Brown & White, 2005, or Hautus & Lee, 2006, for a comparison of different methods and other alternatives). (a) The *Log-linear correction* (Snodgrass & Corwin, 1988) is applied to all frequencies (whatever its value); it is defined as $(H + 0.5) / (N_s + 1)$ for hits and $(F + 0.5) / (N_n + 1)$ for false alarms. (b) *The ± 0.5 correction* (Murdock & Ogilvie, 1968) is applied only if the frequency is 0 (being replaced by 0.5) or N (being replaced by $N - 0.5$), where N is the number of signal or noise trials, as appropriate (alternative values to ± 0.5 have been also proposed for the correction; see Miller, 1996). (c) *Removal* of the proportions that equal 0 and 1 (Miller, 1996). In this procedure the distribution of the proportions of hits and false alarms are truncated, and therefore the distributions of the \hat{z} values associated with such proportions are also truncated. For example, if the procedure of Miller is applied, the summatories appearing in equations [8] and [9] would take values from $i = 1$ to $N_s - 1$, eliminating the addends equal to 0 ($z = -\infty$) and 1 ($z = \infty$).

In sum, the conclusion from many studies has been that the *± 0.5 correction* is the choice proposed for the most common situations. Furthermore, Miller (1996) shows that this correction and the *removal correction* have comparable performance, and better than a correction with a constant less than 0.5. In a Monte Carlo simulation, Hautus (1995) concludes that *log-linear correction* is better than *± 0.5 correction* in order to estimate δ . However, as Kadlec (1999) explains, the simulation conditions used by Hautus are not very realistic (extremely low C criteria and high π_H), being necessary new simulations with more realistic conditions before to accept this conclusion. Moreover, the *± 0.5 correction* uses all data obtained and changing only some of them (in some situations, the probability of applying this is very small). In sum, our choice in this research is the *± 0.5 correction*.

Objectives

In previous studies (e.g., Jesteadt, 2005; Kadlec, 1999; Miller, 1996) several procedures to calculate the variance of d' and c have been compared, but their performance has been assessed by means of the values provided by formulas like [3] – [6]. However, the use of those formulas

requires knowing the parametric values π_H and π_F . The focus in those studies are the estimators d' and c , and how well the cited formulas describe their behavior. On the contrary, we focus here in the estimation of the variance itself, σ_d^2 . In real contexts both π_H and π_F are unknown. Contrary to those previous studies, we focus here in assessing the properties of the estimators of the variance when P_H and P_F replace π_H and π_F . The merits of the three methods are assessed by an evaluation of their bias and precision for a range of values of δ , C , and N .

METHOD

Statistical model

It was assumed the SDT-NH model, with mean 0 and variance 1 for the noise trials and with mean δ (the sensitivity parameter) and variance 1 for the signal trials. Both the frequencies of hits, H , and false alarms, F , were obtained by generating random values. To do that we defined signal and noise distributions, as also the sensitivity parameter, δ . In addition, we set several values for the criterion, C , and the number of signal and noise trials, N_s and N_n . From these values, the probabilities of hits (π_H) and false alarms (π_F) were calculated. Once determined the π_F and N_n values for a given condition, the frequency of false alarms, F , follows a binomial distribution $B(N_n; \pi_F)$ [the frequency of hits, H , follows a binomial distribution $B(N_s; \pi_H)$]. So, the frequencies of hits and false alarms were obtained in the simulation as random values from these distributions.

Conditions of the simulation

Three variables were manipulated: the number of trials of each type, N_s and N_n ; the sensitivity, δ ; and the criterion, C . With respect to the number of trials, both signal and noise always had the same amount: 20, 30, 50, or 80 trials. For δ , the following values were considered: 0.5, 1, 1.5, 2, 2.5, or 3. The criterion values, C , were -0.5, 0, or 0.5. Table 1 shows the values of π_F and π_H corresponding to each pair of values of δ and C . Given the combinations of the levels of the three manipulated variables 72 conditions were simulated (4 numbers of trials x 6 sensitivities x 3 criteria). They were obtained 100,000 repetitions (ie, 100,000 pairs of frequencies of hits and false alarms) for each simulated condition. A program written by the authors in R (R Core Team, 2015) performed the simulation.

Table 1. Values of π_F and π_H used in the simulations. They have been calculated from the values of δ and C , assuming the NH model (see the text).

	C	π_F	π_H
$\delta = 0.5$	-0.5	0.59871	0.77337
	0	0.40129	0.59871
	0.5	0.22663	0.40129
$\delta = 1.0$	-0.5	0.50000	0.84134
	0	0.30854	0.69146
	0.5	0.15866	0.50000
$\delta = 1.5$	-0.5	0.40129	0.89435
	0	0.22663	0.77337
	0.5	0.10565	0.59871
$\delta = 2.0$	-0.5	0.30854	0.93319
	0	0.15866	0.84134
	0.5	0.06681	0.69146
$\delta = 2.5$	-0.5	0.22663	0.95994
	0	0.10565	0.89435
	0.5	0.04006	0.77337
$\delta = 3.0$	-0.5	0.15866	0.97725
	0	0.06681	0.93319
	0.5	0.02275	0.84134

Data analysis

For the pair of frequencies of hits and false alarms of each repetition (H and F) we calculated d' and c assuming the NH model, using equations [1] and [2]. In the event that the frequencies were equal to zero or to the number of trials (N) the ± 0.5 correction was applied (Murdock & Ogilvie, 1968): if the frequency is 0 it is replaced by 0.5 and if the frequency is equal to the number of trials it is replaced by $(N_s - 0.5)$ or $(N_n - 0.5)$. Thus, within each simulated condition 100,000 values of d' and c were obtained. Then, we calculated for each condition the mean and variance of those 100,000 values of d' and c . We checked for departures of those means and variances from the population values. Of course, the population values for the means are the δ values used to generate the data. The population values of variances are those provided by Miller's, formula [5]. We also calculated the population values of the variances of d' by Gourevitch and Galanter's, formula [3].

Table 2 allows assessing the process of data generation by comparing the population value with the means of the d' values. The discrepancies observed in the tables are mainly due to the application of the correction due to zero and N frequencies. Table 3 allows assessing the process of data generation by comparing the population values with the variances of the d' values. It must be remembered that while Miller's method is an exact calculation, Gourevitch and Galanter's method is only an approximation. The discrepancies observed in the variance provided by Miller's formula are mainly due to the application of the correction due to zero and N frequencies. Furthermore, and as was expected, the discrepancy observed in the population value of the variance of d' provided by Gourevitch and Galanter's method are greater than those obtained by Miller's formula, and these depend on δ and the number of trials.

Hereinafter, the estimates of the variance of d' obtained with the three methods set forth in the introduction are compared with the population value obtained with the Miller's method.

Table 2. Means of the 100,000 empirical estimates of d' obtained for each simulated condition.

		N			
C		20	30	50	80
$\delta = 0.5$	-0.5	0.53151	0.51999	0.51151	0.50659
	0	0.52364	0.51229	0.50859	0.50573
	0.5	0.52913	0.52040	0.51082	0.50675
$\delta = 1.0$	-0.5	1.06337	1.04487	1.02485	1.01522
	0	1.04870	1.02963	1.01682	1.01105
	0.5	1.06171	1.04557	1.02662	1.01660
$\delta = 1.5$	-0.5	1.57903	1.57085	1.54670	1.52720
	0	1.58663	1.55545	1.53007	1.51921
	0.5	1.57902	1.57161	1.54502	1.52743
$\delta = 2.0$	-0.5	2.06265	2.08204	2.07389	2.04553
	0	2.12206	2.08785	2.05314	2.02962
	0.5	2.05950	2.08409	2.06884	2.04591
$\delta = 2.5$	-0.5	2.49394	2.55297	2.58188	2.57152
	0	2.63430	2.62635	2.58437	2.54894
	0.5	2.49282	2.55496	2.58047	2.56872
$\delta = 3.0$	-0.5	2.87611	2.96770	3.04586	3.07399
	0	3.07068	3.13005	3.12346	3.08141
	0.5	2.87627	2.97010	3.04313	3.07363

Table 3. Variances of the 100,000 empirical estimates of d' obtained for each simulated condition (*Emp*), variances calculated with Miller's exact method (*M*), and Gourevitch and Galanter's method (*G&G*).

C		N				
		20	30	50	80	
$\delta = 0.5$	-0.5	Emp	0.20487	0.13093	0.07518	0.0458
		M	0.20557	0.13108	0.07503	0.0458
		G&G	0.17698	0.11799	0.07079	0.04425
	0	Emp	0.17848	0.11419	0.06681	0.04076
		M	0.17775	0.11414	0.06666	0.04108
		G&G	0.16069	0.10713	0.06428	0.04017
	0.5	Emp	0.20607	0.13168	0.07399	0.04566
		M	0.20557	0.13108	0.07503	0.0458
		G&G	0.17698	0.11799	0.07079	0.04425
$\delta = 1$	-0.5	Emp	0.22101	0.14809	0.08447	0.05081
		M	0.22057	0.14834	0.08439	0.05072
		G&G	0.19253	0.12835	0.07701	0.04813
	0	Emp	0.19695	0.12426	0.07189	0.04422
		M	0.19799	0.12528	0.0723	0.04432
		G&G	0.17212	0.11475	0.06885	0.04303
	0.5	Emp	0.22182	0.14834	0.08444	0.05076
		M	0.22057	0.14834	0.08439	0.05072
		G&G	0.19253	0.12835	0.07701	0.04813
$\delta = 1.5$	-0.5	Emp	0.22746	0.16987	0.10428	0.06113
		M	0.22538	0.16941	0.10318	0.06101
		G&G	0.22196	0.14798	0.08879	0.05549
	0	Emp	0.23475	0.14838	0.08353	0.05043
		M	0.23338	0.14803	0.0834	0.05052
		G&G	0.19327	0.12885	0.07731	0.04832
	0.5	Emp	0.22544	0.16868	0.10335	0.06084
		M	0.22538	0.16941	0.10318	0.06101
		G&G	0.22196	0.14798	0.08879	0.05549
$\delta = 2$	-0.5	Emp	0.21204	0.1759	0.12645	0.0806
		M	0.21279	0.1764	0.12685	0.08034
		G&G	0.27189	0.18126	0.10875	0.06797
	0	Emp	0.27011	0.18503	0.10414	0.06127
		M	0.26952	0.18589	0.10386	0.06136
		G&G	0.22798	0.15199	0.09119	0.057
	0.5	Emp	0.21383	0.17605	0.12727	0.07967
		M	0.21279	0.1764	0.12685	0.08034
		G&G	0.27189	0.18126	0.10875	0.06797
$\delta = 2.5$	-0.5	Emp	0.19592	0.16502	0.13572	0.10347
		M	0.19515	0.16477	0.13621	0.10388
		G&G	0.35494	0.23662	0.14197	0.08873
	0	Emp	0.27204	0.22346	0.13967	0.0808
		M	0.27301	0.22469	0.1397	0.08094
		G&G	0.28323	0.18882	0.11329	0.07081
	0.5	Emp	0.1974	0.1638	0.13625	0.10354
		M	0.19515	0.16477	0.13621	0.10388
		G&G	0.35494	0.23662	0.14197	0.08873

Table 3 (continued). Variances of the 100,000 empirical estimates of d' obtained for each simulated condition (*Emp*), variances calculated with Miller's exact method (*M*), and Gourevitch and Galanter's method (G&G).

$\delta = 3$	-0.5	Emp	0.18148	0.15244	0.12565	0.10944
		M	0.18147	0.15158	0.1255	0.11043
		G&G	0.49534	0.33022	0.19813	0.12383
	0	Emp	0.22549	0.22835	0.18136	0.11696
		M	0.22759	0.22752	0.1814	0.11636
		G&G	0.37165	0.24777	0.14866	0.09291
	0.5	Emp	0.18157	0.15182	0.12542	0.11103
		M	0.18147	0.15158	0.1255	0.11043
		G&G	0.49534	0.33022	0.19813	0.12383

Within each condition and for each pair of H and F values we obtained estimates of the variance of d' by the three methods set forth in the introduction:

(a) *Method of Gourevitch and Galanter* (1967). Equation [7] was employed for each pair of proportions (P_H and P_F). Thus, for each condition we obtained 100,000 estimates of the variance (100,000 values of $\hat{\sigma}_{d'(GG)}^2$). The mean and the variance of those estimates was calculated for each condition: $\bar{\sigma}_{d'(GG)}^2$ and $S_{\hat{\sigma}_{d'(GG)}^2}^2$.

(b) *Method of Miller* (1996). We repeated the process of the above method but with the equations [8] and [9] (and their counterparts for false alarms), also obtaining 100,000 variance estimates (100,000 values of $\hat{\sigma}_{d'(M)}^2$). Finally, the mean and the variance of the estimates was calculated: $\bar{\sigma}_{d'(M)}^2$ and $S_{\hat{\sigma}_{d'(M)}^2}^2$.

(c) *Method of Dorfman and Alf* (1968). Equation [10] was used as likelihood function for this case, obtaining 100,000 variance estimates (100,000 values of $\hat{\sigma}_{d'(DA)}^2$) and then calculating the mean and the variance: $\bar{\sigma}_{d'(DA)}^2$ and $S_{\hat{\sigma}_{d'(DA)}^2}^2$.

Programs in *R* (R Core Team, 2015) developed by the authors were used for the calculation of d' and c , as well as $\hat{\sigma}_{d'(GG)}^2$, $\hat{\sigma}_{d'(M)}^2$ and $\hat{\sigma}_{d'(DA)}^2$ (see

the appendix). The *bbmle* library (Bolker, 2015) was used for the process of maximum likelihood estimation.

Assessing the performance of the methods of estimation

The bias of the three estimates was assessed by calculating the discrepancy between the population values and the means of the empirical estimates. The bias of an estimate is defined as the difference between the expected value of the estimate and the parameter: $bias = E(\hat{\Theta}) - \Theta$. However, as the importance of the amount of bias must be assessed in relative terms we will calculate the relative bias of the three estimates (Burton, Altman, Royston & Holder, 2006), expressed as a percentage,

$$Relative\ bias = \frac{E(\hat{\Theta}_*) - \Theta}{\Theta} \cdot 100 \quad [11]$$

where $E(\hat{\Theta}_*)$ is the mean estimates of $\sigma_{d'}^2$, computed with each of the three methods ($\bar{\sigma}_{d'(GG)}^2$, $\bar{\sigma}_{d'(M)}^2$ and $\bar{\sigma}_{d'(DA)}^2$) and Θ is the variance of d' obtained with Miller's exact method. A discrepancy close to zero would indicate that the method of estimation is accurate, while positive and negative differences would reflect, respectively, over- and underestimates.

Although the amount of bias must be the main criterion to compare several methods of estimation, it must be complemented with a measure of precision. An unbiased estimate that has a very large variance could be assessed as worse than an estimate with small bias but with a much smaller variance. A good estimate must involve a balanced combination of accuracy and precision. To do that we calculate the mean squared error: $MSE = E(\hat{\Theta}_* - \Theta)^2$; it can be expressed as a function of the bias and the variance of the estimator,

$$MSE = bias^2 + Var(\hat{\Theta}_*) \quad [12]$$

In any practical situation, the researcher has a single estimate of the parameter. Therefore, it is reasonable that the criterion for choosing an estimator be the (squared) expected difference between the estimate and the parameter. This is done by mean of the *MSE*. When comparing the *MSE* values of two competing estimators the amount of bias is penalized by larger variances.

RESULTS AND DISCUSSION

Our main interest is on the ability of the three methods to estimate the actual variance when only the estimates of the probabilities (P_H and P_F) are known. The results for the relative bias are presented in figure 2. This has several striking aspects. First, for all the conditions simulated the relative bias of $\hat{\sigma}_{d(GG)}^2$ and $\hat{\sigma}_{d(DA)}^2$ are virtually identical. In fact, at a first glance the differences are not obvious in the figure because their functions are literally over imposed. Our first conclusion is that, at least for the conditions simulated here, the expected values of both variances are indistinguishable for practical purposes: $\hat{\sigma}_{d(GG)}^2 = \hat{\sigma}_{d(DA)}^2$. Second, in general for the conditions simulated the variance with less relative bias is that obtained with Miller's method ($\hat{\sigma}_{d(M)}^2$); the average of the estimates obtained by this method is closer to the true value than the average obtained by the other two. The larger of these discrepancies is 17.9% (condition with $\delta = 3$, $C = 0$, $N_s = N_n = 30$). Third, the magnitude of the bias with the method of Miller does not change very much across the conditions, and the fluctuations do not show any obvious pattern (they are not systematically associate to δ , C , or N_s and N_n). Fourth, in some conditions the *G&G* and *D&A* methods overestimate considerably the variance in the long run. Those discrepancies increase the higher is δ , the smaller are N_s and N_n , and the farther to 0 is C . In some conditions the relative bias exceeds 140% (for example, the relative bias of the variances estimated by these two methods is 140.5% in the conditions with $\delta = 3$, $C \neq 0$, and $N_s = N_n = 30$).

However, there is a number of conditions for which the amount of bias is not larger for the *G&G* and *D&A* methods than for the Miller's method. See, for example, the conditions with $\delta \leq 1$, or the conditions with $N = 50$ or 80 , with $C = 0$, no matter the value of δ . That is why sometimes has been concluded that there is a range of conditions where those two methods are a reasonable alternative to the Miller's method.

However, a good estimate must have small (if any) bias and large precision (small variance). The *MSE* reflect some balance between both criteria. The results for the *MSE* are presented in figure 3 and table 4. Several aspects must be highlighted also on it. First, the *G&G* and *DA* methods are again practically indistinguishable. Second, the *MSE* for Miller's method outperforms the other two along a range of the conditions simulated, with a few exceptions. In those exceptional occasions (in bold in table 4) the larger *MSE* value for Miller's method is as small as 0.00042 (condition with $\delta = 2$, $C = 0$, $N_s = N_n = 50$).

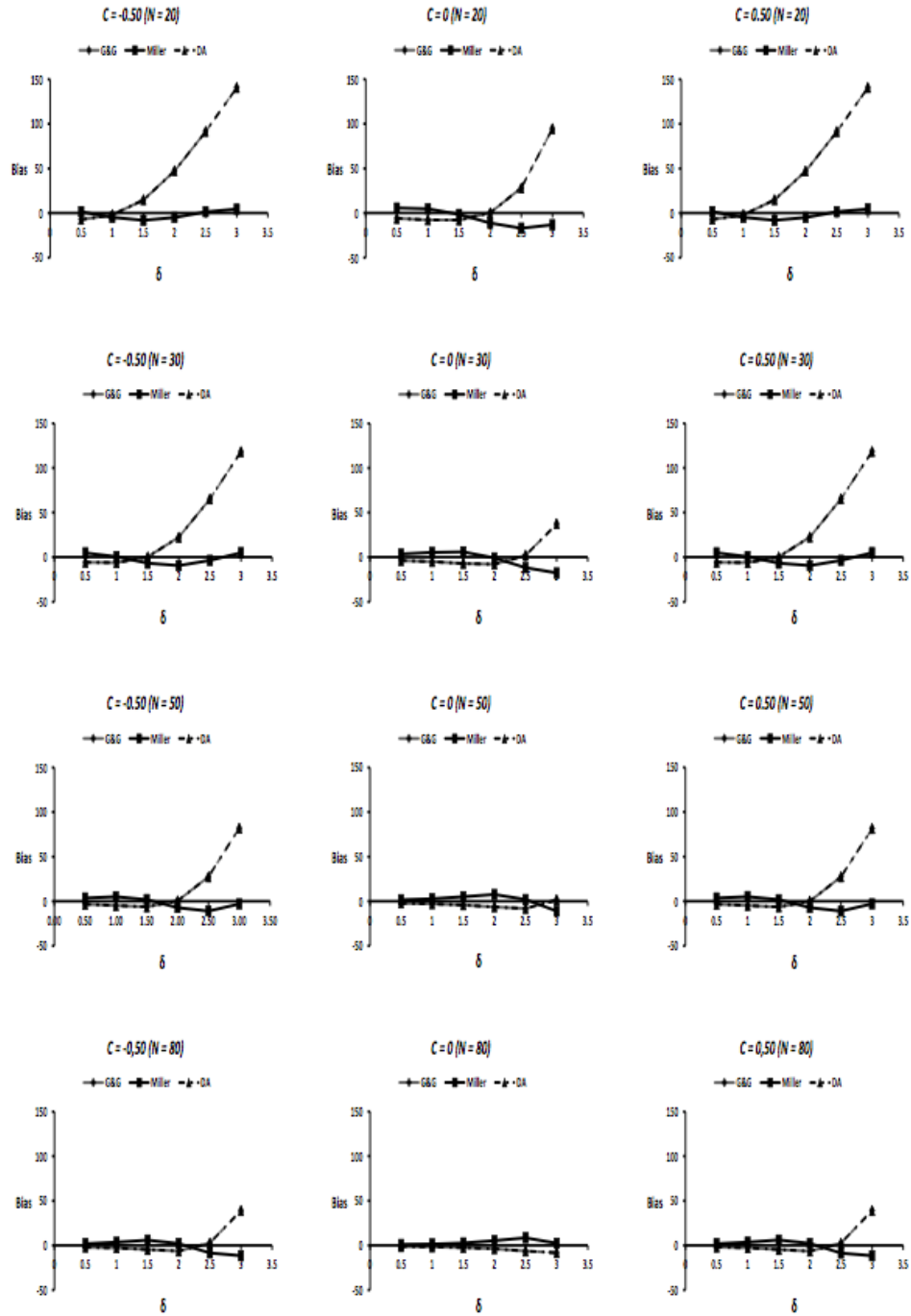


Figure 2. Relative bias (expressed as a percentage) of the three estimation procedures of $\sigma_{d'}^2$ (GG: Gourevitch & Galanter; M: Miller; DA: Dorfman & Alf).

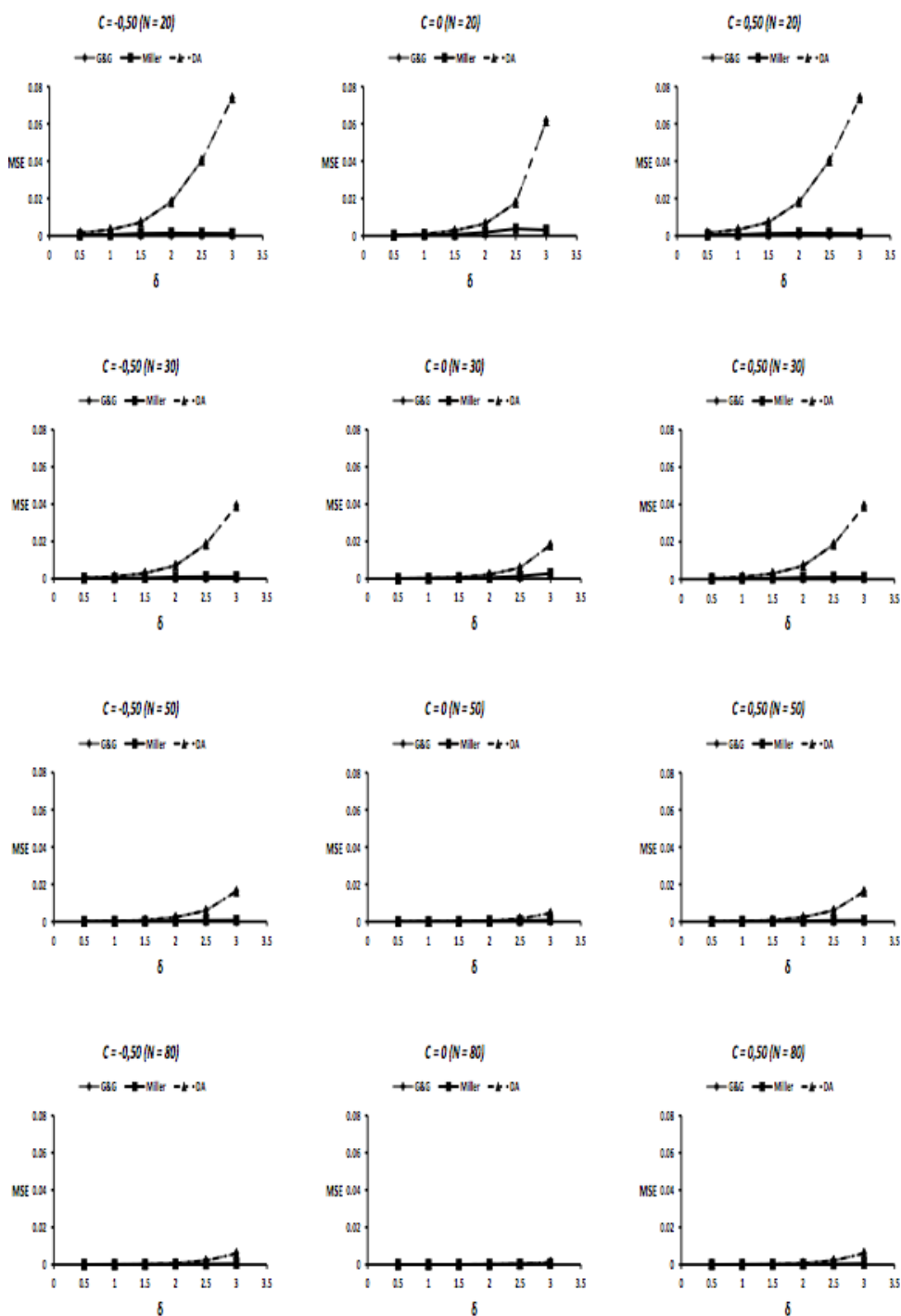


Figure 3. MSE of the three estimation procedures of σ_d^2 . (GG: Gourevitch & Galanter; M: Miller; DA: Dorfman & Alf).

Table 4. Mean squared error of the estimates of the variance of d' obtained with the three methods for each simulated condition.

	C	N															
		20				30				50				80			
		G&G	Miller	DA	DA	G&G	Miller	DA	DA	G&G	Miller	DA	DA	G&G	Miller	DA	DA
$\delta = 0.5$	-0.5	0.00142	0.00041	0.00142	0.00032	0.00031	0.00032	0.00004	0.00004	0.00004	0.00008	0.00004	0.00004	0.00001	0.00001	0.00001	0.00001
	0	0.00027	0.00031	0.00027	0.00004	0.00007	0.00004	0	0.00001	0	0.00001	0	0	0	0	0	0
	0.5	0.00148	0.00041	0.00148	0.00032	0.00030	0.00032	0.00003	0.00003	0.00003	0.00008	0.00003	0.00003	0.00001	0.00001	0.00001	0.00001
$\delta = 1$	-0.5	0.00322	0.00057	0.00322	0.00109	0.00028	0.00109	0.00015	0.00020	0.00015	0.00020	0.00015	0.00002	0.00002	0.00002	0.00002	0.00002
	0	0.00083	0.00057	0.00083	0.00014	0.00028	0.00014	0.00002	0.00004	0.00002	0.00004	0.00002	0	0	0	0	0
	0.5	0.00326	0.00059	0.00326	0.00110	0.00028	0.00110	0.00016	0.00020	0.00016	0.00020	0.00016	0.00002	0.00004	0.00002	0.00002	0.00002
$\delta = 1.5$	-0.5	0.00701	0.00129	0.00701	0.00298	0.00048	0.00298	0.00073	0.00024	0.00073	0.00073	0.00024	0.00011	0.00015	0.00011	0.00011	0.00011
	0	0.00279	0.00062	0.00279	0.00063	0.00055	0.00063	0.00007	0.00015	0.00007	0.00015	0.00007	0.00001	0.00002	0.00001	0.00001	0.00001
	0.5	0.00718	0.00123	0.00718	0.00297	0.00045	0.00297	0.00070	0.00024	0.00070	0.00070	0.00024	0.00011	0.00015	0.00011	0.00011	0.00011
$\delta = 2$	-0.5	0.01827	0.00140	0.01827	0.00710	0.00093	0.00710	0.00230	0.00030	0.00230	0.00230	0.00030	0.00061	0.00021	0.00061	0.00021	0.00061
	0	0.00645	0.00176	0.00645	0.00219	0.00055	0.00219	0.00032	0.00042	0.00032	0.00042	0.00032	0.00005	0.00009	0.00005	0.00005	0.00005
	0.5	0.01787	0.00143	0.01787	0.00711	0.00094	0.00711	0.00227	0.00032	0.00227	0.00227	0.00032	0.00058	0.00021	0.00058	0.00021	0.00058
$\delta = 2.5$	-0.5	0.04014	0.00129	0.04014	0.01845	0.00102	0.01845	0.00596	0.00064	0.00596	0.00596	0.00064	0.00209	0.00025	0.00209	0.00025	0.00209
	0	0.01787	0.00359	0.01787	0.00581	0.00125	0.00581	0.00147	0.00049	0.00147	0.00147	0.00049	0.00023	0.00033	0.00023	0.00023	0.00023
	0.5	0.03949	0.00128	0.03949	0.01877	0.00102	0.01877	0.00596	0.00067	0.00596	0.00596	0.00067	0.00210	0.00026	0.00210	0.00026	0.00210
$\delta = 3$	-0.5	0.07390	0.00117	0.07390	0.03882	0.00096	0.03882	0.01624	0.00070	0.01624	0.01624	0.00070	0.00602	0.00047	0.00602	0.00047	0.00602
	0	0.06251	0.00290	0.06251	0.01792	0.00278	0.01792	0.00452	0.00084	0.00452	0.00452	0.00084	0.00124	0.00041	0.00124	0.00041	0.00124
	0.5	0.07385	0.00118	0.07385	0.03909	0.00096	0.03909	0.01613	0.00069	0.01613	0.01613	0.00069	0.00592	0.00051	0.00592	0.00051	0.00592

Practical implications for meta-analysis

As we noted in the introduction, a meta-analyst usually obtains estimates of effect sizes by a weighted combination of independent estimates of that effect size. The most common weighting scheme is that based on the reciprocal of their variances. When a study reports the mean and variance of the values of d' in two samples of participants the meta-analyst has enough information for applying those procedures. For example, in a study by Rhodes and Jacoby (2007) there are conditions with “frequent” and “infrequent” targets. They report the means and standard deviations of the d' values in the samples. In those cases the sample variance $S_{d'}^2$ can be employed as an estimate of $\sigma_{d'}^2$. However, many papers only report the statistics associated with hits and false alarms rates, and sometimes the values of d' and c associated with the average rates of hits and false alarms. That information could not be enough to obtain the desired estimate of the variance. In this second group of studies $\sigma_{d'}^2$ must be estimated with procedures such as those assessed here. Our results allow us to assess the different alternatives in terms of bias and precision. Many meta-analyses that have been made from the rates of hits and false alarms could be re-done with the statistics d' and c , but this requires to have formulas to calculate estimates of δ and $\sigma_{d'}^2$ from the means and variances of the hits and false alarms rates.

Another problem for the meta-analyst is that the procedures studied here are suitable only if the assumption that all participants in an experimental condition share the same parameter values (δ and C) holds. However, in many situations it is more realistic to assume that there are individual differences in sensitivity and / or criteria among participants of the same experimental condition. To cover this possibility these formulas must be adapted to those situations. We are already working on these new developments (Suero, Botella, & Privado, in preparation).

Among our medium term goals is to develop procedures for meta-analysis of studies within a SDT framework that report partial information. It is very frequent that the studies in several topics only report statistics associated with the rates of hits and false alarms. Consequently, the basis for those meta-analyses are those statistics (e.g., Gardiner, Ramponi, & Richardson-Klavehn, 2002, in recognition memory; Heinrichs & Zakzanis, 1998, in sustained attention). In short, we believe that it is possible to rescue those studies for a meta-analysis based on d' and c statistics and that acknowledges the existence of individual differences in sensitivity and / or

criteria. In that way, we will be able of doing better syntheses of the evidence in topics where SDT is a common framework.

Conclusions

The main conclusion of this study is that among the procedures compared that of Miller (1996) is the most recommended to estimate the variance of d' . In some previous studies it is concluded that in some situations, the *G&G* method is at least equally good, but they are based on the parametric variance (π_F and π_H instead of P_F and P_H) or the methods are assessed only according to their bias. We believe that the methods must be compared assessing both the bias and the *MSE*. When a researcher needs an estimate of the variance of d' what has available are usually P_F and P_H . A good criterion is choosing the estimator for which it is expected a smaller (squared) difference with the population variance: the *MSE*. When *MSE* is taking in account, the recommended estimator must also be Miller's method calculated with the sample proportions. This conclusion is valid for the complete range of conditions assessed in the present study (δ until 3; C between -0.5 and 0.5; N_s and N_n until 80).

All the developments and analyses in this paper refer to data obtained with a Yes/No paradigm. However, our preference for Miller's method converge with the conclusions of simulation studies with rating paradigms (e.g., MacMillan, Rotello, & Miller, 2004). The results of rating experiments allow generating complete ROC curves based in several points in the ROC space. Despite this fundamental difference, the method preferred is the same.

With respect to the variance of the index of response bias, c , as it is based on the same information as d' and this is analyzed in a similar way, the conclusion regarding the estimation methods is the same.

RESUMEN

Métodos para estimar la varianza de algunos índices de la teoría de la detección de señales: Un estudio de simulación. Se presenta un estudio de simulación para evaluar y comparar tres métodos de estimación de la varianza de las estimaciones de los parámetros δ y C de la teoría de la detección de señales (TDS). Se han propuesto varios métodos para calcular la varianza de sus estimadores, d' y c . Dichos métodos han sido evaluados sobre todo comparando las medias y varianzas empíricas en estudios de simulación en los que los cálculos se han hecho con los valores paramétricos de las probabilidades de emitir una respuesta 'si' en un ensayo-señal (aciertos) y en un ensayo-ruido (falsas alarmas). En contextos prácticos la

varianza tiene que ser estimada a partir de las estimaciones de esas probabilidades (tasas empíricas de aciertos y falsas alarmas). Los tres métodos para estimar la varianza comparados en la presente simulación son los basados en la distribución binomial de Miller, en la aproximación a la normal de Gourevitch y Galanter y el de máxima verosimilitud propuesto por Dorfman y Alf. Estos se comparan en términos de su sesgo relativo (exactitud) y en el error cuadrático medio (precisión). Los resultados muestran que los dos últimos métodos se comportan de forma indistinguible a efectos prácticos y producen importantes errores de sobre-estimación en un abanico de situaciones que sin ser las más comunes son bastante realistas en diversos contextos prácticos. Por el contrario, el método de Miller proporciona mejores resultados (o al menos similares) en todas las condiciones estudiadas. Es el método recomendado para obtener estimaciones de las varianzas de estos estadísticos en situaciones aplicadas.

REFERENCES

- Bolker, B. (2015). Package “bbmle”. <http://cran.r-project.org/web/packages/bbmle/bbmle.pdf>
- Borenstein, M., Hedges, L. V., Higgins, J. P. T., & Rothstein, H. R. (2009). *Introduction to meta-analysis*. Chichester, UK: John Wiley and sons.
- Botella, J., & Sánchez-Meca, J. (2015). *Meta-análisis en Ciencias Sociales y de la Salud*. Madrid: Editorial Síntesis.
- Brown, G. S., & White, K. G. (2005). The optimal correction for estimating extreme discriminability. *Behavior Research Methods*, 37(3), 436–449.
- Burton, A., Altman, D. G., Royston, P., & Holder, R. L. (2006). The design of simulation studies in medical statistics. *Statistics in Medicine*, 25, 4279–4292.
- Dorfman, D. D. (1982). RSCORE II. In J. A. Swets & R. M. Pickett (Eds.), *Evaluation of diagnostic systems: Methods from signal detection theory* (pp. 208-232). New York: Academic Press.
- Dorfman, D. D., & Alf, E. (1968). Maximum likelihood estimation of parameters of signal detection theory — A direct solution. *Psychometrika*, 33, 117–124.
- Gardiner, J. M., Ramponi, C., & Richardson-Klavehn, A. (2002). Recognition memory and decision processes: A meta-analysis of remember, know, and guess responses. *Memory*, 10(2), 83-98.
- Gourevitch, V., & Galanter, E. (1967). A significance test for one parameter isosensitivity functions. *Psychometrika*, 32, 25–33.
- Green, D. M., & Swets, J. A. (1966). *Signal detection theory and psychophysics*. New York: Wiley.
- Hautus, M. J. (1995). Corrections for extreme proportions and their biasing effects on estimated values of d' . *Behavior Research Methods, Instruments, & Computers*, 26, 46-51.
- Hautus, M. J., & Lee, A. (2006). Estimating sensitivity and bias in a yes/no task. *The British Journal of Mathematical and Statistical Psychology*, 59 (2), 257–273.
- Hedges, L. V., & Olkin, I. (1985). *Statistical methods for meta-analysis*. Orlando, FL: Academic Press.
- Heinrichs, R. W., & Zakzanis, K. K. (1998). Neurocognitive deficit in schizophrenia: a quantitative review of the evidence. *Neuropsychology*, 12(3), 426.

- Jesteadt, W. (2005). The variance of d' estimates obtained in yes–no and two-interval forced choice procedures. *Perception & psychophysics*, 67(1), 72-80.
- Kadlec, H. (1999). Statistical properties of d' and β estimates of signal detection theory. *Psychological Methods*, 4(1), 22.
- Kaplan, A. (2009). A Comparison of Three Methods for Calculating Confidence Intervals around D-Prime. *Unpublished manuscript*.
- Logan, G. D. (2004). Cumulative Progress in Formal Theories of Attention. *Annual Review of Psychology*, 55, 207-234.
- Macmillan N. A., & Creelman C. D. (2005). *Detection theory: A user's guide*. (2nd ed.). Mahwah, NJ: Lawrence Erlbaum Associates.
- Macmillan, N. A., Rotello, C. M., & Miller, J. O. (2004). The sampling distributions of Gaussian ROC statistics. *Perception & Psychophysics*, 66(3), 406-421.
- Metz, C. E. (1989). Some practical issues of experimental design and data analysis in radiological ROC studies. *Investigative Radiology*, 24, 234-245.
- Miller, J. (1996). The sampling distribution of d' . *Perception & Psychophysics*, 58, 65-72.
- Murdock, B. B., JR., & Ogilvie, J. C. (1968). Binomial variability in short-term memory. *Psychological Bulletin*, 70, 256-260.
- R Core Team (2015). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. <http://www.R-project.org/>.
- Rhodes, M. G., & Jacoby, L. L. (2007). On the dynamic nature of response criterion in recognition memory: Effects of base rate, awareness, and feedback. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 33(2), 305.
- Snodgrass, J. J., & Corwin, J. (1988). Pragmatics of measuring recognition memory: Applications to dementia and amnesia. *Journal of Experimental Psychology: General*, 117, 34–50.
- Suero, M., Botella, J., & Privado, J. (in preparation). Estimating the sampling variance of SDT indexes with heterogeneous individuals.
- Swets, J. A., Dawes, R. M., & Monahan, J. (2000). Psychological Science can improve diagnostic decisions. *Psychological Science in the Public Interest*, 1(1), 1-26.
- Verde, M. F., Macmillan, N. A., & Rotello, C. M. (2006). Measures of sensitivity based on a single hit rate and false alarm rate: The accuracy, precision, and robustness of d' , A_z , and A' . *Perception & psychophysics*, 68(4), 643-654.
- Wickens, T. D. (2001). *Elementary signal detection theory*. Nueva York: Oxford University Press.

APPENDIX**metatds . A R function for computing variance of d' and other indices following three different methods**

```

# COMPUTE:
  #Variance of d' following Gourevitch & Galanter (1967).
  #Mean and variance of d' following Miller (1996).
  #Variance of d' and more (see OUTPUT) following MLE, Dorfman &
  Alf (1968).

# NEEDS PACKAGE: bbmle and stats4.

# ARGUMENTS:
# nr number of noise trials.
# ns number of signal trials.
# pi_fa probability of false alarms or its estimation false alarms rate.
# pi_a probability of hits or its estimation proportions of hits rate.

#OUTPUT is a list with:
# VAR_GG variance d' Gourevitch & Galanter (1967)
# Miller a list with:
  # Varianza variance d' Miller (1996).
  # Val_Esp expected value d' Miller (1996).
# ML a list with:
  #resumen fitting summary
  #p_estim a vector with d' and c estimation
  #loglike is loglike
  #var_covar variance-covariance matrix, var_covar[1,1] is variance
  d'

# Correction extreme values:  $\pm 0.5$  methods. Future version will include
  other methods.

```

```
#####
```

```
library(stats4)
library(bbmle)
```

```
metatds <- function(nr = 100, ns = 100, pi_fa = 0.50, pi_a = 0.50)
{
```

```
#Variance d' Gourevitch & Galanter (1967)
```

```
var_gg <- ((pi_fa*(1-pi_fa))/(nr*dnorm(qnorm(pi_fa))^2)) +
          ((pi_a*(1-pi_a))/(ns*dnorm(qnorm(pi_a))^2))
```

```
#Variance d' and Expected Value d' Miller (1996)
```

```
fre_fa <- c(0.5,(1:(nr-1)), nr-0.5)
fre_a <- c(0.5,(1:(ns-1)), ns-0.5)
```

```
prop_fa <- fre_fa/nr
prop_a <- fre_a/ns
```

```
z_fa <- qnorm(prop_fa, mean = 0, sd = 1)
z_a <- qnorm(prop_a, mean = 0, sd = 1)
```

```
prob_fa <- dbinom(0:nr,nr,pi_fa)
prob_a <- dbinom(0:ns,ns,pi_a)
```

```
v_esp_zfa <- sum(z_fa*prob_fa)
v_esp_za <- sum(z_a*prob_a)
v_esp_miller <- v_esp_za - v_esp_zfa
```

```
var_zfa <- sum(((z_fa*z_fa)*prob_fa))-(v_esp_zfa*v_esp_zfa)
```

```

var_za <- sum(((z_a*z_a)*prob_a))-(v_esp_za*v_esp_za)
var_miller <- var_za + var_zfa

mestim <- list(Val_Esp = v_esp_miller,Varianza=var_miller)

#Variance of d' and more (see OUTPUT) following MLE, Dorfman & Alf
  (1968).
#It is possible other models like normal heteroscedastic or non-normal if LL
is changed.

LL <- function(xc, dp, mr, ms, fal, fac)
-sum((((mr-fal)*pnorm(xc,log.p = TRUE)))+(fal*pnorm(xc,lower.tail =
  FALSE,log.p = TRUE)))+
  (((ms-fac)*pnorm(xc-dp,log.p = TRUE)))+(fac*pnorm(xc-dp,lower.tail =
  FALSE,log.p = TRUE)))

xc_in= qnorm(pi_fa, mean = 0, sd = 1, lower.tail = FALSE)
dp_in= qnorm(pi_a, mean = 0, sd = 1, lower.tail = TRUE) - qnorm(pi_fa,
  mean = 0, sd = 1, lower.tail = TRUE)

fit <- mle2(LL,start=list(xc = xc_in, dp = dp_in) , fixed = list(mr = nr,ms =
  ns, fal= nr*pi_fa, fac = ns*pi_a))

ml <- list(resumen=summary(fit), p_estim=coef(fit, exclude.fixed =
  TRUE),loglike=logLik(fit),var_covar=vcov(fit))

esti <- list(Var_GG =var_gg, Miller=mestim, ML=ml)

return(esti)

}

```