10-2005

The risk of adverse impact in selections based on a test with known effect size

Wilfried DE CORTE
Ghent University

Filip LIEVENS
Singapore Management University, filiplievens@smu.edu.sg
DOI: https://doi.org/10.1177/0013164404272503

THE RISK OF ADVERSE IMPACT IN SELECTIONS BASED ON A TEST WITH KNOWN EFFECT SIZE

WILFRIED DE CORTE
FILIP LIEVENS
Ghent University

The authors derive the exact sampling distribution function of the adverse impact (AI) ratio for single-stage, top-down selections using tests with known effect sizes. Subsequently, it is shown how this distribution function can be used to determine the risk that a future selection decision on the basis of such tests will result in an outcome that reflects the presence of AI. The article therefore provides test and selection practitioners with a valuable tool to decide between alternative selection predictors.

Keywords: *adverse impact; selection; test; effect size*

In both educational and organizational settings, predictors of future achievement such as psychological tests and other measures are often used to decide which candidates will be selected from a given pool of applicants (e.g., Gatewood & Feild, 2001). Typically, candidates are top-down selected, retaining only the higher scoring applicants on a predictor. Although top-down selection on the basis of a valid predictor (i.e., a predictor that correlates positively with the future achievement or criterion behavior) corresponds to the optimal selection rule, the practice may also result in what is commonly referred to as adverse impact (AI). This phenomenon occurs when an applicant group is not homogeneous but instead is a mixture of members from so-called majority and minority populations and when these populations have different average scores on a predictor. In that case, top-down selection is expected to result in different hiring rates for the two candidate groups, with the members of the population with the higher average (i.e., the majority population) being selected more frequently than those of the minority or lower scoring population. The latter finding is AI, and the AI ratio,

Educational and Psychological Measurement, Vol. 63 No. X, Month 2003 1-
DOI: 10.1177/0013164404272503

which is defined as the ratio between the selection rate in the minority applicant group, s_i, and that in the majority group, s_a, is typically used to quantify the extent of the phenomenon: AI Ratio = s_a / s_a (cf. U.S. Equal Employment Opportunity Commission, 1978).

Over the past few decades, the issue of AI has received a lot of attention in the selection literature. The main reason for this is that extensive studies consistently indicate that the most valid predictors of a broad range of academic and organizational criterion behaviors show marked differences in average performance between populations that differ in terms of age, sex, and in particular ethnicity (e.g., Hough, Oswald, & Ployhart, 2001; Sackett, Schmitt, Ellingson, & Kabin, 2001; Schmitt, Clause, & Pulakos, 1996). Thus, using meta-analysis (Hedges & Olkin, 1985; Hunter & Schmidt, 1990) to integrate the results of previous research, it is found that cognitive ability predictors show an effect size (i.e., a standardized mean difference) of about 1 between White and Black populations (Roth, Bevier, Bobko, Switzer, & Tyler, 2001). Data related to the effect size of other important selection predictors, such as personality inventories, biodata questionnaires, and selection interviews, have become available as well (e.g., Bobko, Roth, & Potosky, 1999; Hough et al., 2001; Ones & Anderson, 2001).

In conjunction with the above-reported studies, other research has focused on the expected level of AI as a function of the effect size of the intended selection predictors (e.g., Sackett & Ellingson, 1997; Schmitt, Rogers, Chan, Sheppard, & Jennings, 1997). To study this relationship, it is assumed that the predictor scores have a normal distribution with the same variance but a different mean value in the majority and the minority applicant populations. As indicated by previous research (e.g., Crawford, Gray, & Allan, 1995; Plomin, 1999; Schmidt, Hunter, McKenzie, & Muldrow, 1979; Tiffin & Vincent, 1960), the assumption is often adequate for general ability predictors in unscreened candidate populations. Equating, without a loss of generality, the within-population variance of the predictor scores to 1, and using Y and X to denote the predictor scores in the minority and the majority applicant populations, respectively, this assumption can be rewritten as $Y \sim N(0, 1)$ and $X \sim N(\delta, 1)$, where δ is the population effect size of the predictor. A selection with an overall selection rate, S, from a total applicant population with mixture proportions π_a and $\pi_i = 1 - \pi_a$ for the majority and the minority applicants is then expected to result in an AI ratio equal to $[1 - \Phi(p_c)] / [1 - \Phi(p_c - \delta)]$, where $\Phi(\bullet)$ denotes the standard normal distribution function, and p_c is the predictor cutoff value such that the intended overall selection ratio is achieved: $\pi[1 - \Phi(p_c)] + \pi_a[1 - \Phi(p_c - \delta)] = S$ (e.g., Morris, 2001; Sackett & Ellingson, 1997).

The above procedure to assess the extent of AI shows two major problems. First, the procedure assumes that the applicant group is infinitely large and hence results in the population value of the AI ratio. However, selections always relate to finite applicant samples, and it is shown hereafter that this population value often offers only a crude estimate of the level of AI one may really expect to obtain with small or medium-sized applicant samples. Second, the approach, henceforth referred to as the population approach, is inadequate to clarify the way in which the actual value of the AI ratio will vary from one applicant sample (or from one application of the selection predictor) to another. To address both problems, this article details the exact sampling distribution function (SDF) of the AI ratio, given the number of applicants from the majority and the minority populations (n_a and n_i, respectively); the number of selected applicants, m; and the value of the population effect size δ of the intended selection predictor. Under the above-detailed assumption that $Y \sim N(0, 1)$ and $X \sim N(\delta, 1)$, and assuming that the applicant pool is a random sample from the total population, the SDF tabulates the exact cumulative probabilities for the different possible sample AI ratio values as a function of the predictor effect size parameter δ. Observe that the dependency on δ is shared by the above-discussed population method to assess the expected AI. Also, although the value of δ is typically unknown, this should not pose a major problem, because fairly accurate estimates of the population effect size are available for many currently used selection predictors (cf. the above-cited studies). The practical importance of the SDF thus obtained is further amplified by showing how this distribution can subsequently be used to assess the probability that an as yet unimplemented selection will result in a selection outcome that reflects the condition of AI (see below). The latter probability will henceforth be referred to as the risk of AI.

As discussed below, alternative formulations of the SDF of the AI ratio are easily obtained using, for example, results on the distribution of the ratio of two probabilities (cf. Agresti, 2002). Unfortunately, these alternative expressions cannot immediately be applied to study the sampling variability of the AI ratio for an intended but not yet implemented selection because these expressions depend on the population minority and majority selection rates for which, in that case, no estimates are available. To resolve this problem, an explicit assumption must be made with respect to the distribution of the predictor scores in the two applicant populations such that the required estimates can be inferred from the population effect size value of the intended selection predictor, the number of minority and majority candidates in the applicant pool, and the required number of selectees. The distributions thus obtained, although now also expressed in terms of the predictor population effect size, will not match the presently derived SDF of the AI ratio, however. As a consequence, the assessment of the risk of AI, as computed from these distributions, will differ from the correct assessment, as based on the present exact SDF, and it will be illustrated that the discrepancy may often be not only of theoretical but of practical importance as well.

The next section discusses the derivation of the exact (i.e., small-sample) SDF of the sample AI ratio as a function of the population effect size δ of the selection predictor, given the number of minority and majority candidates in the applicant pool and the required number of selectees. Following this, it is shown how this SDF can be used to assess the risk of AI of an intended selection. The relationship between the present proposal and the available statistical tests to assess AI is explored in yet a further section. Other potential applications of the approach, as well as its limitations and possible extensions, are discussed in the final two sections of the article.

SDF and the Expected Value of the AI Ratio

The presently derived SDF of the AI ratio applies to the same situations as those addressed by the above-described population approach in that both developments focus on selections in which candidates are top-down selected on the basis of their scores on the predictor. The derivation of the SDF relies also on the same stochastic model for the predictor scores as the one invoked to determine the population AI. Thus, it is throughout understood that $Y \sim N(0, 1)$ in the minority applicant population and that $X \sim N(\delta, 1)$ in the majority applicant population, with δ the given population effect size of the predictor. However, in contrast to the population derivation, knowledge of the mixing proportions π_i and π_a is no longer assumed, and the reference to the population overall selection rate, S, is dropped as well. Instead, the derivation is conditional on the numbers n_i and n_a of candidates from the minority and the majority applicant populations in the actual applicant sample. Both numbers, together with the number of candidates that one intends to select, m, determine the sample overall selection rate, s, and the numbers n_i, n_a, and m are henceforth considered as given, fixed quantities. Finally, as is standard practice (e.g., Stuart & Ord, 1994), it is assumed that the actual applicants represent random draws from their respective populations. Observe that the latter assumption is in fact less restrictive than the corresponding assumption of the population approach, whereby the applicant sample is effectively equated to the applicant population. Also, as discussed in the final section, the random-sample assumption can, if required, be replaced by a more suitable one.

To obtain the SDF of the sample AI ratio A, henceforth denoted as $F(A)$, it is first observed that this function is completely specified by the distribution function of the number selected from the minority applicant group. In general, the latter distribution function is defined for values, j, of the number selected from the minority group, J, that are in the range $l \leq j \leq u$, where $l = \max(0, m - n_a)$ and $u = \min(m, n_i)$. Also, with top-down selection, results from the theory of order statistics (David, 1981) can be used to determine the values of the distribution function of J. To apply these results, let $Y_{(j)}$ indicate the jth-order statistic from the sample of minority group observations; that is,

$Y_{(j)}$ corresponds to the jth smallest predictor score in the sample of the minority candidates, such that $Y_{(l)}$ and $Y_{(n_i)}$ represent the smallest and the largest predictor score within this sample, respectively. Similarly, let $X_{(i)}$ denote the ith-order statistic of the sample of majority group predictor scores. The probability that at most j (with $l \le j \le u$) candidates from the minority group are selected, $P(J \le j)$, can then be equated to the probability that $Y_{(n_i - j)}$ has a smaller value than $X_{(n_a - (m-j)+1)}$. Next, because of the independence of the observations in the two samples, it is further obtained that

$$P(J \le j) = P\left\{X_{[n_a - (m-j)+1]} > Y_{[n_i - j]}\right\}$$
$$= \int_{-\infty}^{+\infty}\int_{-\infty}^{x} f(x)\, g(y)\, dy\, dx,$$

where $f(x)$ and $g(y)$ represent the density functions of the order statistics $X_{(n_a - (m-j)+1)}$ and $Y_{(n_i - j)}$, respectively. Using the earlier discussed distributional assumptions, these densities can subsequently be detailed as

$$f(x) = \frac{n_a!}{\left[n_a - (m-j)\right]!\,(m-1-j)!}$$
$$\times\, \phi(x-\delta) \times \left[\Phi(x-\delta)\right]^{n_a - (m-j)} \times \left[1 - \Phi(x-\delta)\right]^{m-1-j},$$

and

$$g(y) = \frac{n_i!}{(n_i - j - 1)!\,(j)!} \times \phi(y) \times \left[\Phi(y)\right]^{n_i - j - 1} \times \left[1 - \Phi(y)\right]^{j},$$

where $\phi(\bullet)$ and $\Phi(\bullet)$ indicate the standard normal density and distribution function, respectively. Obviously, for $j = u$, $P(J \le j = u) = 1$, and the SDF of the AI ratio, $F(A)$, for values $a_j = (j / n_i) / [(m - j) / n\text{va}]$ with $l \le j \le u \ne m$ can be obtained by equating $F(A = a_j)$ to $P(J \le j)$. Also, for $j = u = m$, the value of the AI ratio is not defined, but the corresponding value $F(A = a_m)$ can in that case still be conveniently equated to 1.

The above-detailed derivation of $F(A)$ shows that the determination of the expected value of the AI ratio, $E(A)$, is straightforward only in the case that $m > n_i$. Otherwise, there is a nonzero probability that none of the candidates from the majority applicant group is selected, resulting in an undefined expected value. To avoid the latter complication it is suggested that the determination of $E(A)$ be limited to only the finite terms of a_j. More specifically, and using $f(a_j)$ to denote the probability mass function of the AI ratio, $E(A)$ is henceforth defined as

$$E(A) = \sum_{j=l}^{m-1} f(a_j)\, a_j \Big/ \sum_{j=1}^{m-1} f(a_j),$$

when $m \leq n_i$ and as

$$E(A) = \sum_{j=l}^{u1} f(a_j) a_j,$$

otherwise. This "unorthodox" proposal is of little practical consequence, however, unless the required number of selected applicants is very small (i.e., less than one or two). In all other cases, the probability of the event that only minority applicants are selected can be virtually neglected because the majority applicant population typically outperforms the minority applicant population on the predictors that are generally used in actual selection applications. Also, as discussed below, the proposal does not affect the assessment of the risk that a future selection will result in an AI outcome.

When the population effect size of the selection predictor, δ, equals zero, the SDF of the AI ratio reduces, for $l \leq j < u$, to

$$F(A = a_j) = \sum_{k=0}^{j} \frac{m-j}{m-k} \binom{n_a}{m-j} \binom{n_i}{j-k} \binom{n}{m-k},$$

where $n = n_i + n_a$ denotes the total number of applicants. Through computation, it can be further verified that the above expression equals

$$\sum_{k=l}^{j} \binom{n_i}{k} \binom{n-n_i}{m-k} \binom{n}{m} = \sum_{k=l}^{k} h(k; m, n_1, n),$$

with $h(k; m, n_i, n)$ the familiar hypergeometric distribution with parameters m, n_i, and n. This is a somewhat unexpected result, given the difference between the above-detailed assumptions and the stochastic model that underlies the hypergeometric distribution.

With the above result in mind, it might be suggested that the SDF $F(A)$ can in general (i.e., for effect size values that differ from zero) be modeled as an extended (cumulative) hypergeometric distribution. More specifically, it might be proposed that for $\delta \neq 0$, $F(A)$ can be written as

$$\sum_{k=l}^{j} h(k; m, n_1, n, \theta),$$

where

$$h(k; m, n_1, n, \theta) = \frac{\binom{n_i}{k} \binom{n-n_i}{m-k} \theta^k}{\sum_{l \leq t \leq u} \binom{n_i}{t} \binom{n-n_i}{m-t} \theta^t}$$

is the extended hypergeometric distribution (Harkness, 1965; Johnson, Kotz, & Kemp, 1993) and expresses the ratio of the odds that a candidate from the minority applicant population will be selected and the corresponding odds for a candidate from the majority population. Also, given the present assumptions (i.e., $Y \sim N[0, 1]$ and $X \sim N(\delta, 1]$), the latter odds ratio can be further specified as

$$\theta = \frac{\left[1 - \Phi\left(p_c\right)\right]\Phi\left(p_c - \delta\right)}{\Phi\left(p_c\right)\left[1 - \Phi\left(p_c - \delta\right)\right]},$$

where p_c is as defined above.

However, it is easily verified that the thus proposed formulation does not match the exact SDF $F(A)$. The discrepancy between the two expressions can be explained by observing that the extended hypergeometric formulation is based on the assumption that the distribution of the number selected from the minority applicant group, J, can be modeled as the conditional distribution of the first of two binomial random variables, given that their sum is fixed (cf. Johnson et al., 1993). The two binomial variables involved are the number of minority selected applicants, J (with $J \sim \text{Bin}[n_i, q_i]$ and q_i the selection probability of a minority applicant) and the number of selected majority applicants, K (with $K \sim \text{Bin}[n_a, q_a]$), whereas the fixed-sum condition is $J + K = m$, the total number selected. This representation fails in the present context, however. Although it is reasonable to assume that under random sampling, the selection probabilities q_a and q_i are constant within each total applicant sample of size $n_i + n_a$, this is no longer true over different total applicant samples. Over the latter samples, the values of q_a and q_i must be allowed to vary to account for the fact that the selection probability of a minority (majority) applicant depends on the set of actually sampled values of the $n_i + n_a$ predictor scores. Thus, the q_i value is expected to be larger (smaller) in a total applicant sample where the majority applicants happen to have rather low (high) predictor scores. Finally, observe that the above variability will be of no effect (i.e., the effect will cancel out evenly over repeated samples) in case that the population effect size of the predictor, δ, equals zero, which explains why the hypergeometric distribution fits the presently derived SDF of the AI ratio for $\delta = 0$.

Illustration

The following two examples illustrate the above detailed results. The first example presents the SDF of the AI ratio, $F(A)$, associated with a selection using predictors that vary in their population effect size value. The second example focuses on the difference between the expected value of the AI ratio,

$E(A)$, as obtained in the present exact (i.e., finite-sample) approach and the corresponding value predicted by the traditional population method.

To illustrate the determination of $F(A)$, consider the situation in which students, 240 from the majority (i.e., $n_a = 240$) and 60 from the minority population (i.e., $n_i = 60$), apply for one of the 30 available openings (i.e., $m = 30$) in a medical education program. In that case, 31 different selection outcomes and hence 31 different values for the AI ratio are possible. Either $j = 0, 1, 2, \ldots, 30$ minority candidates, with corresponding numbers selected from the majority group of $m - j = 30, 29, \ldots, 0$, can be selected. Also, the associated set of possible values for the AI ratio is, for $0 \le j \le m - 1$, $\{a_0, a_1, \ldots, a_{m-1}\} = \{0.000, 0.138, \ldots, 116.000\}$; whereas for $j = m$, the value of the corresponding AI ratio, a_m, is undefined.

Although the above example selection situation is always characterized by the same set of possible values for the AI ratio, regardless of the value of the population effect size of the selection predictor, δ, it is obvious that the probability with which these values will occur over repeated samples of candidates will vary as a function of δ. This is shown in Table 1, which details (part of) the SDF of the AI ratio for the example selection when predictors with values of $\delta = 0.0$, 0.2, 0.5, and 1.0 are used to perform the selection. By and large, the values of δ are chosen to reflect rather realistic selection situations. Thus, the choice $\delta = 1.0$ corresponds to the use of a college application test such as the SAT in a situation in which the applicants belong to different ethnic populations (cf. Roth et al., 2001). Alternatively, $\delta = 0.0$ indicates the use of a neutral predictor or the implementation of random selection.

For each value of the AI ratio, a_j, the tabled values correspond to the probability that this value or a smaller value will be obtained. Thus, for a neutral predictor it is shown that the probability to obtain a selection outcome with an associated value of, for example, at most 0.615 for the AI ratio (i.e., the value of the SDF evaluated at $a_j = 0.615$, $F[A = a_j = 0.615]$) is 0.242. The corresponding probability when using a test with an effect size of 0.5 is 0.885, whereas the use of a predictor with an effect size of 1.0 is almost certain to result (i.e., with .998 probability) in an AI ratio that is not larger than 0.615. Alternatively, the Table 1 values are also useful to derive an (approximate) 90% probability interval for the sample AI ratio given the value of δ. For the present example this interval extends from 0.286 to 1.714 for $\delta = 0.0$, whereas the interval is bounded by 0.000 and 0.800 in case that $\delta = 0.5$.

To illustrate that the traditional population approach results in a biased estimate of the expected AI ratio value, $E(A)$, compared with the corresponding correct value obtained by the present exact method, the second example explores the difference between the two resulting values of $E(A)$ for a variety of selection scenarios. To exclude the above-discussed eventuality that the value of $E(A)$ may be undefined, only scenarios for which the number selected, m, is larger than the number of minority group applicants, n_i, are

Table 1
Sampling Distribution Function of the Adverse Impact (AI) Ratio When Selecting 30 Candidates From a Total of 300 Applicants (60 minority and 240 majority candidates), Using a Selection Test With Population Effect Size, δ, Equal to 0.0, 0.2, 0.5, and 1.0

			Effect Size Selection Test			
J	*K*	AI Ratio	δ = 0.0	δ = 0.2	δ = 0.5	δ = 1.0
0	30	0.000	0.001	0.007	0.058	0.394
1	29	0.138	0.008	0.044	0.237	0.770
2	28	0.286	0.037	0.146	0.495	0.940
3	27	0.444	0.110	0.321	0.732	0.988
4	26	0.615	0.242	0.532	0.885	0.998
5	25	0.800	0.420	0.725	0.960	1.000
6	24	1.000	0.609	0.862	0.988	1.000
7	23	1.217	0.770	0.942	0.997	1.000
8	22	1.455	0.883	0.979	0.999	1.000
9	21	1.714	0.949	0.994	1.000	1.000
10	20	2.000	0.981	0.998	1.000	1.000
11	19	2.316	0.994	1.000	1.000	1.000
12	18	2.667	0.998	1.000	1.000	1.000
13	17	3.059	1.000	1.000	1.000	1.000

Note. J indicates the number of selected minority applicants. *K* indicates the number of selected majority applicants.

considered. More specifically, the studied situations vary in terms of (a) the overall selection rate, *s*, with $s = 0.15$, 0.25, and 0.50; (b) the total number of applicants, *n*, with $n = 20$, 40, 100, and 500; and (c) the population value of the selection predictor, δ, with $\delta = 0.0$, 0.2, 0.5, and 1.0. Also, all situations are characterized by the same ratio, equal to 0.9, between the number of majority candidates and the number of minority applicants.

Table 2 summarizes the results of the above-detailed comparison. In general, these results confirm that the population method (cf. the values reported in the rows for which the total number of candidates is ∞) underestimates the $E(A)$ value for all finite samples and hence for all realistic selections.

The underestimation does not relate to the normal distribution assumption for the predictor scores, because this assumption is shared by both the present exact and the former population methods. Also, the underestimation is especially important for small sample selections with a low selection rate. Thus, for selections with a selection rate *s* of, for example, 0.15, a total number of applicants of 20, and a selection predictor effect size of 0.50, the underestimation (compared with the exact value) equals as much as $(0.641 - 0.421) \times 100 / 0.641 = 34.3\%$. For larger applicant samples, the results between the traditional and the present exact method tend to converge, however, but even for medium-sized applicant groups (e.g., $n = 100$) there remains a noticeable

Table 2
Expected Adverse Impact Ratio, E(A)*, for Selections That Vary in Terms of the Overall Selection Rate, the Total Number of Applicants, and the Population Effect Size of the Selection Predictor, Given an Applicant Group Consisting of 90% Majority and 10% Minority Applicants*

	Number of	Effect Size Selection Predictor			
Selection Ratio	Applicants	δ = 0.0	δ = 0.2	δ = 0.5	δ = 1.0
0.15	20	1.492	1.084	0.641	0.232
0.15	40	1.191	0.864	0.508	0.181
0.15	100	1.068	0.773	0.452	0.159
0.15	500	1.013	0.733	0.427	0.148
0.15	Infinite	1.000	0.723	0.421	0.146
0.25	20	1.204	0.929	0.600	0.252
0.25	40	1.092	0.839	0.538	0.222
0.25	100	1.035	0.793	0.507	0.208
0.25	500	1.007	0.771	0.492	0.200
0.25	Infinite	1.000	0.766	0.488	0.199
0.50	20	1.092	0.895	0.635	0.314
0.50	40	1.044	0.853	0.602	0.295
0.50	100	1.017	0.830	0.585	0.285
0.50	500	1.003	0.818	0.576	0.280
0.50	Infinite	1.000	0.815	0.574	0.279

difference between the population-based value and the corresponding correct value.

Further Comments

Although the population method underestimates the AI ratio value that one may expect to obtain, the above results show that the underestimation need not always be of practical importance. In addition, the refinement in the determination of $E(A)$ as obtained by the present method can be questioned by observing that many selections are part of an ongoing program, resulting in a total applicant sample (as aggregated over the years) that is quite large. As an example, consider a university that each year uses the same predictor (or composite predictor) to admit candidates to a Ph.D. program. In such a case, and because of the larger aggregate sample, the population and the present methods will produce similar values of $E(A)$. Observe, however, that this aggregation of small sample selections to one large sample selection is appropriate only in the case that an identical predictor (or predictor composite) is used at each instance. In particular, the use of the same predictors, but weighed somewhat differently over the years, and changing the predictor basis to include an additional predictor are both practices that preclude the aggregation of the selections. Such practices are no exception, though, as organizations continuously strive for more optimal selection strategies. In addition, many selections are not part of ongoing programs but occur on an ad hoc basis to meet such irregular demands for new employees as implied by, for example, a start-up or a reallocation of business and production activities.

Whereas the present refinement in the estimation of $E(A)$ is less practically relevant for large sample selections, this is much less the case as far as the derivation of the SDF of the AI ratio is concerned. Obviously, the latter distribution will become more narrowly centered for larger applicant samples, but the SDF will continue to show considerable variability unless the sample is extremely large. To illustrate this, consider the selection of 750 new employees from an applicant group of 5,000 candidates. With 4,000 majority and 1,000 minority applicants, and using a predictor with an effect size of, for example, 0.4, the (approximate) 90% probability interval for the AI ratio will in that case still extend from 0.425 (5.7 percentile) to 0.608 (95.3 percentile). This shows that the present derivation of the SDF of the AI ratio remains of substantial importance, even for large sample selection decisions. In addition, and as discussed in the next section, this SDF can also be used to asses the risk that an intended selection, using a predictor with a known effect size, will result in an AI outcome.

Assessing the Risk of AI

To assess the probability that a future application of a selection predictor will lead to an AI outcome, two related issues must be resolved. First, it must be decided which of the possible selection outcomes will be judged to reflect AI. Second, one must determine the probability that these AI outcomes will be obtained when the predictor is applied to perform a future selection. As shown hereafter, the first issue is resolved either by convention or through the use of a statistical test, whereas the present derivation of the SDF of the AI ratio addresses the second issue.

Determination of the AI Outcomes

Given the total set of possible selection outcomes as determined by the values of n_a, n_i and m, the first task is to assess which, if any, of these outcomes will be considered as unacceptable because they are characterized by too few minority hires compared with the corresponding number of majority hires. Although the distinction between AI and no AI outcomes is intrinsically judgmental and therefore cannot be solved in any absolute way, two approaches are currently used to settle the issue (e.g., Morris, 2001). In the first, convention-based approach, the so-called four-fifths rule of thumb

(U.S. Equal Employment Opportunity Commission, 1978) is invoked to distinguish between unacceptable (AI) and acceptable (no AI) outcomes. According to this rule, a selection outcome reflects AI if the selection rate for the minority group is less than four fifths of the selection rate for the majority. In the second approach, a statistical test, such as Fisher's exact test, is used to determine if an outcome is unacceptable.

Obviously, both approaches frequently lead to a different demarcation of AI versus no-AI outcomes. To illustrate this, reconsider the above-discussed example situation in which 30 candidates are selected from a candidate sample that consists of 240 applicants from the majority population and 60 candidates from the minority population. In this situation, and as can be verified from the data in Table 1, only the outcomes for which at most 4 minority applicants are selected will be judged to reflect AI when the four-fifths rule is used to decide which outcomes show AI. The application of Fisher's exact test does not lead to the same conclusion, however, because the corresponding hypergeometric distribution function (which, as argued above, is identical to the presently derived SDF of the AI ratio for $\delta = 0$) indicates that the probability to select 4 or fewer minority candidates is equal to .242 (cf. the second column of Table 1). This is a much higher value than the .05 probability that corresponds to a one-sided test at the conventional 5% significance level. Using the latter significance level, the hypergeometric distribution probabilities and hence Fisher's exact test show that only the outcomes for which no more than 2 minority applicants are selected are unacceptable.

The above-exemplified discrepancy between the results of the two approaches to determine AI outcomes does not affect the below-detailed procedure to assess the risk of AI, however. As argued above, the distinction between AI and no AI outcomes is essentially judgmental and the risk assessment can be performed irrespective of the approach used to distinguish between the two types of outcome.

Determination of the Risk of an AI Outcome

Given the adopted distinction between AI and no-AI outcomes, the next step is to assess the risk (i.e., the probability) that a future selection on the basis of a predictor with known population effect size will lead to one of the AI outcomes. This assessment requires the specification of the SDF of the selection outcomes that corresponds to the population effect size of the selection predictor. However, the latter SDF is identical to the SDF of the AI ratio statistic obtained in the previous section because each possible value of the AI ratio corresponds to exactly one of the different possible selection outcomes. As a consequence, the presently derived SDF of the AI ratio suffices to assess the risk of an AI outcome.

To illustrate the procedure, consider again the example situation in which 30 candidates are selected from an applicant pool with 240 majority and 60 minority candidates. For this situation, the application of Fisher's exact test indicated that only the outcomes with at most 2 minority admissions are unacceptable. Suppose now that the organization intends to use a predictor that has a population effect size value, δ, equal to 0.5. In that case, the SDF of the AI ratio that corresponds to this effect size shows a probability of 0.495 (the value in the third row of the sixth column in Table 1) and therefore a risk of nearly 50% that the latter selection will result in an AI outcome (i.e., an outcome for which fewer than 3 minority candidates are admitted). Alternatively, if admissions were based on a predictor with an effect size value of 0.2, the results in Table 2 indicate a much lower risk of only 14.6% of obtaining an AI outcome.

Relationship Between Selection Characteristics and Risk of AI

To provide a more general overview of the relationship between the characteristics of the intended selection and the corresponding risk of obtaining an AI selection outcome, the risk is determined for a variety of intended selection decisions. The studied selections differ in terms of (a) the proportional representation of the majority and minority candidates in the total applicant sample (i.e., either 90% vs. 10% or 80% vs. 20%), (b) the overall selection rate, s (i.e., $s = 0.05$, 0.10, 0.25, and 0.50), (c) the total number of candidates, n (i.e., $n = 40$, 100, 200, 300, 500, and 1,000), and (d) the effect size δ of the intended selection predictor (i.e., $\delta = 0.10$, 0.20, 0.30, 0.50, and 0.80). Table 3 summarizes the obtained results. For each selection, two values are reported for the risk that the intended selection will result in an AI outcome. The first value is calculated from the exact SDF of the AI ratio, whereas the second value, in parentheses, corresponds to an alternative estimate of the risk of AI. As detailed in the next section, the alternative estimate is based on a large-sample approximation of the SDF of the AI ratio.

Apart from the obvious result that predictors with higher effect size values result in higher risk values, the tabled values also show that selections with as many as 300 to 500 applicants may have an associated risk of AI equal to zero, especially when the selection ratio is low and the proportional representation of the minority applicants is small. The reason for this is that in these cases, Fisher's exact test does not permit one to distinguish between AI and no-AI outcomes, because the probability that none of the minority candidates is selected using a neutral predictor exceeds the conventional 5% significance level. The results furthermore indicate that the risk of AI is higher for higher selection rates, larger total numbers of applicants, and a more substantial proportional representation of the minority candidates in the applicant

(text continues on p. #)

Table 3

Risk of Adverse Impact (AI) of an Intended Selection as Related to the Proportional Representation of Minority Applicants, the Overall Selection Rate, the Total Number of Applicants, and the Predictor Effect Size

	Number of	Effect Size Selection Predictor				
Selection Ratio	Applicants	δ = 0.1	δ = 0.2	δ = 0.3	δ = 0.5	δ = 0.8
10% minority candidates in the applicant sample						
0.05	40	0.00 (**)	0.00 (**)	0.00 (**)	0.00 (**)	0.00 (**)
0.05	100	0.00 (**)	0.00 (**)	0.00 (**)	0.00 (**)	0.00 (**)
0.05	200	0.00 (**)	0.00 (**)	0.00 (**)	0.00 (**)	0.00 (**)
0.05	300	0.00 (**)	0.00 (**)	0.00 (**)	0.00 (**)	0.00 (**)
0.05	500	0.00 (**)	0.00 (**)	0.00 (**)	0.00 (**)	0.00 (**)
0.05	1,000	0.07 (0.14)	0.13 (0.28)	0.23 (0.43)	0.47 (0.69)	0.80 (0.85)
0.10	40	0.00 (**)	0.00 (**)	0.00 (**)	0.00 (**)	0.00 (**)
0.10	100	0.00 (**)	0.00 (**)	0.00 (**)	0.00 (**)	0.00 (**)
0.10	200	0.00 (**)	0.00 (**)	0.00 (**)	0.00 (**)	0.00 (**)
0.10	300	0.06 (**)	0.10 (**)	0.15 (**)	0.28 (**)	0.52 (**)
0.10	500	0.06 (0.13)	0.11 (0.24)	0.18 (0.38)	0.39 (0.63)	0.72 (0.82)
0.10	1,000	0.13 (0.16)	0.26 (0.34)	0.44 (0.54)	0.78 (0.82)	0.98 (0.95)
0.25	40	0.00 (**)	0.00 (**)	0.00 (**)	0.00 (**)	0.00 (**)
0.25	100	0.07 (**)	0.10 (**)	0.14 (**)	0.24 (**)	0.43 (**)
0.25	200	0.04 (0.11)	0.07 (0.19)	0.11 (0.30)	0.24 (0.53)	0.52 (0.77)
0.25	300	0.07 (0.12)	0.12 (0.23)	0.21 (0.37)	0.44 (0.64)	0.80 (0.86)
0.25	500	0.09 (0.14)	0.20 (0.30)	0.35 (0.48)	0.69 (0.78)	0.96 (0.95)
0.25	1,000	0.11 (0.19)	0.30 (0.44)	0.55 (0.69)	0.92 (0.94)	1.00 (1.00)
0.50	40	0.00 (**)	0.00 (**)	0.00 (**)	0.00 (**)	0.00 (**)
0.50	100	0.07 (0.10)	0.11 (0.16)	0.17 (0.24)	0.31 (0.43)	0.57 (0.68)
0.50	200	0.09 (0.11)	0.16 (0.21)	0.26 (0.33)	0.51 (0.59)	0.84 (0.85)

0.50	300	0.09 (0.13)	0.18 (0.25)	0.31 (0.41)	0.62 (0.71)	0.94 (0.93)
0.50	500	0.08 (0.15)	0.19 (0.33)	0.36 (0.54)	0.76 (0.86)	0.99 (0.99)
0.50	1,000	0.15 (0.21)	0.39 (0.49)	0.68 (0.76)	0.98 (0.98)	1.00 (1.00)
20% minority candidates in the applicant sample						
0.05	40	0.00 (**)	0.00 (**)	0.00 (**)	0.00 (**)	0.00 (**)
0.05	100	0.00 (**)	0.00 (**)	0.00 (**)	0.00 (**)	0.00 (**)
0.05	200	0.00 (**)	0.00 (**)	0.00 (**)	0.00 (**)	0.00 (**)
0.05	300	0.06 (**)	0.10 (**)	0.15 (**)	0.30 (**)	0.56 (**)
0.05	500	0.06 (0.13)	0.11 (0.25)	0.19 (0.40)	0.42 (0.66)	0.76 (0.84)
0.05	1,000	0.13 (0.17)	0.27 (0.36)	0.47 (0.57)	0.82 (0.85)	0.99 (0.96)
0.10	40	0.00 (**)	0.00 (**)	0.00 (**)	0.00 (**)	0.00 (**)
0.10	100	0.00 (**)	0.00 (**)	0.00 (**)	0.00 (**)	0.00 (**)
0.10	200	0.02 (**)	0.04 (**)	0.06 (**)	0.15 (**)	0.37 (**)
0.10	300	0.08 (0.12)	0.15 (0.24)	0.24 (0.38)	0.50 (0.65)	0.83 (0.86)
0.10	500	0.10 (0.15)	0.22 (0.31)	0.38 (0.50)	0.73 (0.79)	0.97 (0.95)
0.10	1,000	0.15 (0.20)	0.36 (0.46)	0.62 (0.71)	0.95 (0.95)	1.00 (1.00)
0.25	40	0.00 (**)	0.00 (**)	0.00 (**)	0.00 (**)	0.00 (**)
0.25	100	0.03 (0.10)	0.05 (0.18)	0.08 (0.27)	0.20 (0.49)	0.45 (0.74)
0.25	200	0.07 (0.12)	0.13 (0.24)	0.23 (0.39)	0.51 (0.68)	0.87 (0.91)
0.25	300	0.08 (0.14)	0.18 (0.30)	0.33 (0.49)	0.70 (0.80)	0.97 (0.97)
0.25	500	0.14 (0.18)	0.33 (0.40)	0.58 (0.64)	0.92 (0.93)	1.00 (1.00)
0.25	1,000	0.20 (0.25)	0.52 (0.60)	0.83 (0.86)	1.00 (0.99)	1.00 (1.00)
0.50	40	0.03 (**)	0.05 (**)	0.08 (**)	0.15 (**)	0.33 (**)
0.50	100	0.07 (0.10)	0.13 (0.19)	0.21 (0.30)	0.43 (0.55)	0.78 (0.82)
0.50	200	0.07 (0.13)	0.15 (0.26)	0.27 (0.44)	0.61 (0.75)	0.95 (0.96)
0.50	300	0.09 (0.15)	0.22 (0.33)	0.41 (0.55)	0.81 (0.87)	0.99 (0.99)
0.50	500	0.17 (0.19)	0.40 (0.45)	0.68 (0.71)	0.97 (0.97)	1.00 (1.00)
0.50	1,000	0.26 (0.28)	0.64 (0.67)	0.91 (0.92)	1.00 (1.00)	1.00 (1.00)

Note. The first value in each cell corresponds to the risk based on the exact sampling distribution function of the AI ratio. The second value, reported in parentheses, is the corresponding risk based on the large-sample approximate distribution of $\log(A)$. The latter risk is indicated by two asterisks ("**") in case the large-sample approximation is not applicable because the expected number of selectees from the smallest applicant group is less than five.

sample, but these main effects must be further qualified because they also tend to interact. It therefore seems less advisable to propose a set of simple rules to gauge the risk of AI as related to the characteristics of the intended selection. Instead, the results suggest that the practitioner can better apply the present risk calculation using the actual values of the number of majority and minority candidates, the number of required selectees, and the effect size estimate of the intended selection predictor.

Relationship With Traditional Statistical Tests for Adverse Impact

Thus far, the discussion has focused on the assessment of the risk that an intended selection will result in a selection outcome that reflects AI. Although this focus differs from the one adopted in traditional statistical tests (e.g., the *z*-test of Morris & Lobsenz, 2000) of AI, these tests are often based on distributions of the sample AI ratio (or a simple transform of this ratio) that can also lead to such a risk assessment. The resulting assessment will be inaccurate, however, because these distributions either are less appropriate, as is the case for the above-discussed extended hypergeometric distribution, or represent large-sample approximations of the exact distribution. Also, these distributions typically depend on population values for the minority and majority selection rates instead of on the predictor population effect size value δ such that their application to assess the risk of AI of a not yet implemented selection requires a transformation of the available sample and predictor information. But this transformation—and it is important to emphasize this—cannot succeed unless an explicit assumption is made with respect to the distribution of the predictor scores in the majority and the minority candidate populations. Alternative approaches to the risk assessment of a future selection, as based on these other distributions, therefore require the same assumptions as the present proposal. In particular, both the present and the alternative approaches need to assume that the majority and the minority candidates represent random draws from their respective populations, and they all depend on an explicit model as to the way in which the predictor scores are distributed in these populations.

To demonstrate the above conclusion, as well as to show that the alternative approaches may often result in poor risk estimates, the *z*-test proposed by Morris and Lobsenz (2000) is considered in some detail. This *z*-test is based on the finding that the natural log of the sample AI ratio A, $\log(A)$, is approximately normally distributed when the applicant sample is large (Agresti, 2002), with expectation equal to $\log \alpha$, where $\alpha = S_i / S_a$ indicates the population AI ratio, and S_i and S_a correspond to the minority and the majority population selection rates. Also, the standard deviation of the distribution, $\sigma_{\log(A)}$, is

$$\sigma_{\log(A)} = \sqrt{\frac{1-S_i}{n_i S_i} + \frac{1-S_a}{n_a S_a}},$$

where n_i and n_a represent, as before, the number of minority and majority candidates in the applicant sample (cf. Morris & Lobsenz, 2000).

To apply the above distribution to assess the risk of AI, the population selection rates S_i and S_a must be estimated. Yet, given the available selection information (i.e., the values of n_i, n_a, and m, as well as an estimate of δ), it is obvious that this estimation requires an explicit assumption with respect to the distribution of the predictor scores in the two applicant populations. Thus, using as before the assumption that $Y \sim N(0, 1)$ and $X \sim N(\delta, 1)$, the population selection rates S_i and S_a can be determined as $1 - \Phi(p_c)$ and $1 - \Phi(p_c - \delta)$, respectively, where the predictor cutoff value p_c is such that the intended overall selection rate s, with $s = m / (n_i + n_a)$, is achieved given the actual proportional representation of the minority and minority candidates in the sample applicant group:

$$s = \frac{m}{n_i - n_a} = \frac{n_i}{n_i + n_a}\left[1-\Phi\left(p_c\right)\right] + \frac{n_a}{n_i + n_a}\left[1-\Phi\left(p_c - \delta\right)\right].$$

Because the z-test relies on approximate, large-sample distributions, the related assessment of the risk of AI will be credible only in case that the normal approximation is sufficiently acceptable. This means in particular that the usage of these distributions (and of the z-test in particular) is recommended only for selection situations in which the expected number of selected candidates from the smallest applicant group is at least equal to five (cf. Morris & Lobsenz, 2000). Yet, even for selections that meet this requirement, the resulting risk assessment may still differ from the correct assessment to such a degree that the difference is practically meaningful as well. To illustrate this, consider a situation in which 400 majority and 100 minority candidates compete for 25 available vacancies, such that the expected number of selected minority applicants equals 5, and the large-sample approximate distribution of $\log(A)$ can be applied. With random selection (i.e., selection based on a neutral predictor with population effect size δ equal to zero), it is then obtained that the distribution of $\log(A)$ is approximately normal, with an expected value of zero and a standard deviation of

$$\sqrt{\frac{1-.05}{100 \times .05} + \frac{1-.05}{400 \times .05}} = 0.487.$$

Using the conventional 5% significance level, the one-sided z-test therefore indicates that selection outcomes with an associated value of $\log(A)$ of at most $-1.645 \times 0.487 = -0.801$ will be considered as outcomes that reflect AI. Suppose now that instead of random selection, one intends to use a predictor with a population effect size equal to 0.2. In that case, the assumption on the normal distribution of the predictor scores, together with the available values of $n_i = 100$, $n_a = 400$, and $m = 25$, permits one to determine the majority and minority selection rates that correspond to this effect size value as 0.054 and 0.035, respectively. Thus, when the effect size equals 0.2, log(A) has an expected value of log(0.035/0.054) = –0.423 and a standard deviation of

$$\sqrt{\frac{1-.035}{100\times.035}+\frac{1-.054}{400\times.054}}=0.565..$$

The risk to obtain an AI outcome, using the predictor with effect size 0.2, therefore corresponds to the probability that a normal distributed random variable with expectation –0.423 and standard deviation 0.565 will not exceed the earlier determined critical value of –0.801. The latter probability equals .25, and it can be verified from the results in Table 3 that this result is more than twice the correct risk value of .11, as determined on the basis of the exact SDF of the AI ratio. Also, given the large discrepancy between the two risk assessments, it seems difficult to deny that the difference is also practically relevant.

Obviously, the discrepancy between the correct and the above derived, large sample based risk assessment will become smaller when the exact SDF converges to the corresponding large-sample approximation. As observed above, this convergence depends on the expected number of selected applicants from the smallest applicant group, but this general rule does not indicate how the two risk estimates actually compare for a broad range of selection situations. We therefore computed the risk, according to the large-sample approximation of the SDF of the AI ratio, for the entire set of earlier analyzed selection situations and added the results in parentheses following the corresponding correct values already reported in Table 3. For selection situations with an expected number of selected minority applicants that is less than five, the large-sample approximation of the SDF of the AI ratio is generally not recommended (see above), and the associated risk value is therefore reported as two asterisks ("**"). In general, the results confirm the earlier finding that assessments based on approximate sampling distributions are not trustworthy unless the total applicant sample is large. Given this observation, and the fact that the approximate assessment requires both identical assumptions and identical data as the exact procedure, the usage of the former assessment can only be justified on the grounds that it is simpler to compute. However, even this final justification no longer applies because an easily applicable computer program to perform the correct risk calculation is made available to the interested audience. The program, as well as two additional codes to handle the below discussed extensions, can be downloaded at http://allserv.rug.ac.be/~wdecorte/software.html.

Some Extensions

Thus far, the derivation of the exact SDF of the AI ratio and the subsequent assessment of the risk of AI has been obtained in the understanding that the composition of the applicant pool, in terms of the numbers of majority and minority candidates, is available. This may not always be the case, however, in that the group membership of the applicants may become known only at the time that the selection is actually performed. But even then the present results can still be used to derive the SDF of the AI ratio. To obtain this distribution, it is then assumed, just as in the traditional procedure, that a population estimate of the proportion of minority applicants in the total candidate population is available. The assumption implies that the number of minority applicants in the applicant pool has a binomial distribution: $n_i \sim \mathrm{Bin}(n, \pi_i)$, with n the total number of applicants and π_i the proportion of minority candidates in the total applicant population. Using $b(ni)$ to denote the corresponding binomial frequency function, and letting $F(A \mid n_i)$ denote the SDF of the AI ratio, given that the number of minority applicants equals n_i, it then follows that the (marginal) SDF function of the AI statistic is equal to

$$\sum_{n_i=1}^{n_i=n-1} b(n_i) F(A|n_i) \Big/ \sum_{n_i=1}^{n_i=n-1} b(n_i) .$$

The latter (marginal) distribution function of the AI ratio can subsequently be used as before to assess the risk of obtaining an AI outcome.

A second extension addresses an issue that can be particularly relevant with very small sample selection decisions. For such selections, it may be unrealistic to assume that the minority and majority applicants are random representatives of their respective populations, because the actual applicant pool is the result of some systematic recruitment effort. One way to address this situation is to consider the actual candidate pool not as a random sample of the initial population but rather as a sample from a restricted subset of this population and to adapt the predictor distribution assumption accordingly. To illustrate the proposal, consider an example in which the recruited candidates are prescreened on the basis of rudimentary biographical information. In that case, the rate of selectivity applied in the prescreen, together with an estimate of the correlation between the biographical prescreen and the intended selection predictor, can be combined with standard results on truncated binormal distributions to derive a more suitable expression for the distribution of the predictor scores in the prescreened applicant populations. Using the latter distribution instead of the normal distribution, the earlier discussed procedure can then again be applied to obtain the SDF of the AI ratio and to assess the risk of AI. Obviously, the above proposal is not relevant when the required information is not available. However, in that case, all procedures, including the population approach, for the prediction of AI will fail.

Discussion

In the previous sections of this article, we presented and exemplified a two-step procedure that enables selection practitioners to estimate the risk that a future selection decision will lead to an outcome that reflects AI. When compared with other procedures that serve a similar purpose, it was found that the present proposal is always to be preferred because it is based on the exact instead of an approximate or an inadequate formulation of the SDF of the AI ratio and because it achieves the risk estimation without any additional assumptions. In this section, we briefly discuss the dependency of the procedure on the population predictor effect size value and indicate a further extension of our method. We also delineate the limitations of the procedure.

Just as the traditional procedure to estimate the expected AI, the present method requires data on the population effect size δ of the intended selection predictor. It could therefore be argued that both procedures are rather pointless because, as shown above, knowledge of δ implies also knowledge of the value of the AI ratio at the population level. The argument is not appropriate, however, in that it is shown above that even a predictor with a substantial population effect size may, when applied to a given sample of candidates, result in a selection outcome that does not show AI. AI at the population level is quite different from AI at the actual sample level. Selection practitioners always deal with applicant samples, never with the applicant population. They are therefore best helped by a procedure, such as the present one, that focuses on the selection results that may be expected at the sample level. Obviously, the accuracy of the results thus obtained will depend on the precision of the population effect size estimate, but it is repeated that such fairly accurate estimates have become available for most of the popular selection predictors (e.g., Bobko et al., 1999; Hough et al., 2001; Roth et al., 2001). Also, to account for the remaining imprecision of these estimates, the risk assessment can be determined for both the lower- and the upper-bound values of the predictor effect size, resulting in an optimistic as well as a conservative assessment of the risk. Finally, observe that the dependency of the risk assessment on the predictor effect size value is entirely similar to the way in which, for example, the expected quality of an intended selection, as expressed in terms of the expected criterion performance of the selected candidates, depends on the estimated (population) validity of the predictor. Predictions of future selection outcomes necessarily rely on estimated quantities. If the latter estimates are reasonably trustworthy, as can be argued for both validity and effect size estimates, it seems good practice to make optimal use of this information, because otherwise, a selection practitioner is left without guidance when deciding between alternative predictors.

Thus far, this article has focused on selections in which a single predictor is used to perform the selection, but selection decisions are often based on the aggregate result (i.e., the weighed sum) of the scores as obtained on several different predictors. The extension of the present proposal to selections on the basis of such aggregate or composite predictors is straightforward, however, provided that estimates of the predictor correlations and effect sizes are available. For in that case, the effect size of the composite, δ_c, can be simply determined as $\delta_c = \mathbf{b}'\delta \,/\, \mathbf{b}'\mathbf{R}\mathbf{b}$, where δ is the vector of individual predictor effect sizes, **b** represents the weights with which the individual predictors are combined to the composite predictor, and **R** denotes the correlation matrix of the predictors. With this extension, the risk assessment can be performed not only for single-predictor selections but for composite-predictor selections as well.

Even with the above extension, the procedure still shows a number of limitations. As recognized from the onset, the method can be applied only to single-stage selection situations, and it is highly unlikely that this limitation can be overcome. Although the approach can, at least in principle, be extended to cope with multistage situations, the resulting expressions for the SDF of the AI statistic are numerically intractable, leaving no other option than to use Monte Carlo simulation methods to study the sampling variability of the statistic for other than single-stage selections. Also, even for single-stage decisions, it must be remembered that the results of the procedure are valid provided that the underlying stochastic model is appropriate and that the predictor data that are used to determine the composite effect size are accurate. The latter dependency is not unique to the present proposal, however, because the currently used method to derive the population estimate of the AI statistic is based on even stronger assumptions and requires identical predictor data to obtain the estimate. For single-stage decisions, it can therefore be concluded that the present procedure provides selection practitioners with a valuable tool to decide between alternative selection predictors. The proposal may also contribute to the construction of predictor composites for situations in which both the goals of diversity and quality are of importance. Because of this potential and as the program to implement the method is made widely available, it is hoped that the procedure will find routine application in the design of selection decisions in both educational and organizational settings.

References

Agresti, A. (2002). *Categorical data analysis* (2nd ed.). New York: John Wiley.

Bobko, P., Roth, P. L., & Potosky, D. (1999). Derivation and implications of a meta-analytic matrix incorporating cognitive ability, alternative predictors, and job performance. *Personnel Psychology*, *52*, 561-589.

Crawford, J. R., Gray, C. D., & Allan, K. M. (1995). The WAIS-R (UK) basic psychometric properties in an adult UK sample. *British Journal of Clinical Psychology*, *34*, 237-250.

David, H. A. (1981). *Order statistics* (2nd ed.). New York: John Wiley.

Gatewood, R. D., & Feild, H. S. (2001). *Human resource selection*. Fort Worth, TX: Harcourt College.

Harkness, W. L. (1965). Properties of the extended hypergeometric distribution. *Annals of Mathematical Statistics*, *36*, 938-945.

Hedges, L. V., & Olkin, I. (1985). *Statistical methods for meta-analysis*. London: Academic Press.

Hough, L. M., Oswald, F. L., & Ployhart, R. E. (2001). Determinants, detection and amelioration of adverse impact in personnel selection procedures: Issues, evidence and lessons learned. *International Journal of Selection and Assessment*, *9*, 152-194.

Hunter, J., & Schmidt, F. (1990). *Methods of meta-analysis: Correcting error and bias in research findings*. Beverly Hills, CA: Sage.

Johnson, N. L., Kotz, S., & Kemp, A. W. (1993). *Univariate discrete distributions*. New York: John Wiley.

Morris, S. B. (2001). Sample size requirements for adverse impact analysis. *Applied HRM Research*, *6*, 13-32.

Morris, S. B., & Lobsenz, R. E. (2000). Significance tests and confidence intervals for the adverse impact ratio. *Personnel Psychology*, *53*, 89-111.

Ones, D. S., & Anderson, N. (2002). Gender and ethnic group differences on personality scales in selection: Some British data. *Journal of Occupational and Organizational Psychology*, *75*, 255-276.

Plomin, R. (1999). Genetics and general cognitive ability. *Nature*, *402*, C25-C29.

Roth, P. I., Bevier, C. A., Bobko, P., Switzer, F. S., III, & Tyler, P. (2001). Ethnic group differences in cognitive ability in employment and educational settings: a meta-analysis. *Personnel Psychology*, *54*, 297-330.

Sackett, P. R., & Ellingson, J. E. (1997). The effects of forming multi-predictor composites on group differences and adverse impact. *Personnel Psychology*, *50*, 707-721.

Sackett, P. R., Schmitt, N., Ellingson, J. E., & Kabin, M. B. (2001). High-stakes testing in employment, credentialing and higher education. *American Psychologist*, *56*, 302-318.

Schmidt, F. L., Hunter, J. E., McKenzie, R. C., & Muldrow, T. W. (1979). Impact of valid selection procedures on workforce productivity. *Journal of Applied Psychology*, *64*, 609-626.

Schmitt, N., Clause, C. S., & Pulakos, E. D. (1996). Subgroup differences associated with different measures of some common job-relevant constructs. In C. L. Cooper & I. T. Robertson (Eds.), *International review of industrial and organizational psychology* (Vol. 11, pp. 115-139). New York: John Wiley.

Schmitt, N., Rogers, W., Chan, D., Sheppard, L., & Jennings, D. (1997). Adverse impact and predictive efficiency of various predictor combinations. *Journal of Applied Psychology*, *82*, 719-730.

Stuart, A., & Ord, K. J. (1994). *Kendall's advanced theory of statistics: Volume 1, Distribution theory* (6th ed.). London: Edward Arnold.

Tiffin, J., & Vincent, N. L. (1960). Comparison of empirical and theoretical expectancies. *Personnel Psychology*, *13*, 59-64.

U.S. Equal Employment Opportunity Commission. (1978). Uniform guidelines on employment selection procedures. *Federal Register*, *43*, 38290-38315.